(72) Inventors; and
(75) Inventors/Applicants (for US only): KAZEMI, Amir [US/US]; 53 Southwind Drive, Wallingford, CT 06492 (US). KLIEM, Stefanie, E. [DE/US]; Kiefernweg 37, 61440 Oberusel (DE). LANZ, Elizabeth, M. [US/US]; 111 Park Street, No. 5U, New Haven, CT 06511 (US). MESSER, Chad [US/US]; 41 Manor Street, Hamden, CT 06517 (US). TANGUAY, Debra, A. [US/US]; 902 Aspen Glen Drive, Hamden, CT 06518 (US).

(74) Agents: FIELD, Gisela, M. et al.; Genaissance Pharma-ceuticals, Inc., Five Science Park, New Haven, CT 06511 (US).

(54) Title: HAPLOTYPES OF THE POR GENE

(57) Abstract: Novel genetic variants of the P450 (Cytochrome) Oxidoreductase (POR) gene are described. Various genotypes, haplotypes, and haplotype pairs that exist in the general United States population are disclosed for the POR gene. Compositions and methods for haplotyping and/or genotyping the POR gene in an individual are also disclosed. Polynucleotides defined by the haplotypes disclosed herein are also described.

HAPLOTYPES OF THE POR GENE

RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Application Serial 60/236,449 filed
5   September 29, 2000.

FIELD OF THE INVENTION

This invention relates to variation in genes that encode pharmaceutically-important proteins.
In particular, this invention provides genetic variants of the human P450 (cytochrome) oxidoreductase
10   (POR) gene and methods for identifying which variant(s) of this gene is/are possessed by an
individual.

BACKGROUND OF THE INVENTION

Current methods for identifying pharmaceuticals to treat disease often start by identifying,
15   cloning, and expressing an important target protein related to the disease. A determination of whether
an agonist or antagonist is needed to produce an effect that may benefit a patient with the disease is
then made. Then, vast numbers of compounds are screened against the target protein to find new
potential drugs. The desired outcome of this process is a lead compound that is specific for the target,
thereby reducing the incidence of the undesired side effects usually caused by activity at non-intended
20   targets. The lead compound identified in this screening process then undergoes further *in vitro* and *in
vivo* testing to determine its absorption, disposition, metabolism and toxicological profiles. Typically,
this testing involves use of cell lines and animal models with limited, if any, genetic diversity.

What this approach fails to consider, however, is that natural genetic variability exists between
individuals in any and every population with respect to pharmaceutically-important proteins, including
25   the protein targets of candidate drugs, the enzymes that metabolize these drugs and the proteins whose
activity is modulated by such drug targets. Subtle alteration(s) in the primary nucleotide sequence of a
gene encoding a pharmaceutically-important protein may be manifested as significant variation in
expression, structure and/or function of the protein. Such alterations may explain the relatively high
degree of uncertainty inherent in the treatment of individuals with a drug whose design is based upon a
30   single representative example of the target or enzyme(s) involved in metabolizing the drug. For
example, it is well-established that some drugs frequently have lower efficacy in some individuals
than others, which means such individuals and their physicians must weigh the possible benefit of a
larger dosage against a greater risk of side effects. Also, there is significant variation in how well
people metabolize drugs and other exogenous chemicals, resulting in substantial interindividual
35   variation in the toxicity and/or efficacy of such exogenous substances (Evans et al., 1999, *Science*
286:487-491). This variability in efficacy or toxicity of a drug in genetically-diverse patients makes
many drugs ineffective or even dangerous in certain groups of the population, leading to the failure of

1

such drugs in clinical trials or their early withdrawal from the market even though they could be highly beneficial for other groups in the population. This problem significantly increases the time and cost of drug discovery and development, which is a matter of great public concern.

It is well-recognized by pharmaceutical scientists that considering the impact of the genetic
5   variability of pharmaceutically-important proteins in the early phases of drug discovery and development is likely to reduce the failure rate of candidate and approved drugs (Marshall A 1997 *Nature Biotech* 15:1249-52; Kleyn PW et al. 1998 *Science* 281: 1820-21; Kola I 1999 *Curr Opin Biotech* 10:589-92; Hill AVS et al. 1999 in *Evolution in Health and Disease* Stearns SS (Ed.) Oxford University Press, New York, pp 62-76; Meyer U.A. 1999 in *Evolution in Health and Disease* Stearns
10   SS (Ed.) Oxford University Press, New York, pp 41-49; Kalow W et al. 1999 *Clin. Pharm. Therap.* 66:445-7; Marshall, E 1999 *Science* 284:406-7; Judson R et al. 2000 *Pharmacogenomics* 1:1-12; Roses AD 2000 *Nature* 405:857-65). However, in practice this has been difficult to do, in large part because of the time and cost required for discovering the amount of genetic variation that exists in the population (Chakravarti A 1998 *Nature Genet* 19:216-7; Wang DG et al 1998 *Science* 280:1077-82;
15   Chakravarti A 1999 *Nat Genet* 21:56-60 (suppl); Stephens JC 1999 *Mol. Diagnosis* 4:309-317; Kwok PY and Gu S 1999 *Mol. Med. Today* 5:538-43; Davidson S 2000 *Nature Biotech* 18:1134-5).

The standard for measuring genetic variation among individuals is the haplotype, which is the ordered combination of polymorphisms in the sequence of each form of a gene that exists in the population. Because haplotypes represent the variation across each form of a gene, they provide a
20   more accurate and reliable measurement of genetic variation than individual polymorphisms. For example, while specific variations in gene sequences have been associated with a particular phenotype such as disease susceptibility (Roses AD *supra*; Ulbrecht M et al. 2000 *Am J Respir Crit Care Med* 161: 469-74) and drug response (Wolfe CR et al. 2000 *BMJ* 320:987-90; Dahl BS 1997 *Acta Psychiatr Scand* 96 (Suppl 391): 14-21), in many other cases an individual polymorphism may be found in a
25   variety of genomic backgrounds, i.e., different haplotypes, and therefore shows no definitive coupling between the polymorphism and the causative site for the phenotype (Clark AG et al. 1998 *Am J Hum Genet* 63:595-612; Ulbrecht M et al. 2000 *supra*; Drysdale et al. 2000 *PNAS* 97:10483-10488). Thus, there is an unmet need in the pharmaceutical industry for information on what haplotypes exist in the population for pharmaceutically-important genes. Such haplotype information would be useful in
30   improving the efficiency and output of several steps in the drug discovery and development process, including target validation, identifying lead compounds, and early phase clinical trials (Marshall et al., *supra*).

One pharmaceutically-important gene for the treatment of cancers is the P450 (cytochrome) oxidoreductase (POR) gene or its encoded product. POR is a flavoprotein localized to the nuclear
35   envelope and endoplasmic reticulum of most cell types. POR is responsible for the transfer of electrons from NADPH to cytochrome P450 as well as heme oxygenase, squalene epoxidase and cytochrome b(5) (O' Leary KA and Kasper CB. *Arch Biochem Biophys* 2000 379:97-108).

Cytochrome P450-based cancer gene therapy is a novel prodrug activation strategy for cancer treatment. The primary goal of this type of therapy is to selectively increase tumor cell exposure to cytotoxic drug metabolites generated locally by a prodrug-activating P450 enzyme. This leads to a more efficient killing of the transduced tumor cells without a significant increase in host cytotoxicity

5   (Waxman et al., *Drug. Metab. Rev.* 1999 May:503-22).   An example of this type of therapy is illustrated by the use of Tirapazamine (TPZ), an anticancer drug that is activated to a cytotoxic radical intermediate in low oxygen environments, as occur in cancer cells. Cancer cells have reduced activity of POR, and, consequently, these cells are more resistant to TPZ under both aerobic and hypoxic conditions when compared to normal cells. When transfected with POR, however, cancer cell

10   sensitivity to TPZ increases under both aerobic and hypoxic conditions (Saunders et al., *Br. J. Cancer* 2000.82:651-6). It is well known that TPZ is particularly effective when combined with alkylating agents such as cyclophosphamide (CPA). In a study by Jounaidi Y. and Waxman DJ (*Cancer Res.* 2000 60:3761-9), the effects of TPZ with CPA in a cytochrome P450 based prodrug activation gene therapy were examined. Gliosarcoma cells were transduced with genes encoding cytochrome P450

15   2B6 and with POR. TPZ exhibited a striking growth-inhibitory effect on the CPA-treated cytochromeP450 2B6/POR tumor cells under both normoxic and hypoxic conditions compared to tumor cells treated with TPZ alone or in combination with CPA. These results clearly show the utility of the cytochrome P450 based prodrug activation strategy in the effective treatment of cancer.

        The P450 (cytochrome) oxidoreductase gene is located on chromosome 7q11.2 and contains

20   16 exons that encode a 676 amino acid protein. A reference sequence for the POR gene is shown in the contiguous lines of Figure 1(Genaissance Reference No. 6118814; SEQ ID NO: 1). Reference sequences for the coding sequence (GenBank Accession No. S90469.1) and protein are shown in Figures 2 (SEQ ID NO: 2) and 3 (SEQ ID NO: 3), respectively.

        Because of the potential for variation in the POR gene to affect the expression and function of

25   the encoded protein, it would be useful to know whether polymorphisms exist in the POR gene, as well as how such polymorphisms are combined in different copies of the gene. Such information could be applied for studying the biological function of POR as well as in identifying drugs targeting this protein for the treatment of disorders related to its abnormal expression or function.

30   SUMMARY OF THE INVENTION

        Accordingly, the inventors herein have discovered 26 novel polymorphic sites in the POR gene. These polymorphic sites (PS) correspond to the following nucleotide positions in Figure 1: 22635 (PS1), 29577 (PS2), 29754 (PS3), 30427 (PS4), 31531 (PS5), 31644 (PS6), 31648 (PS7), 32446 (PS8), 32478 (PS9), 33538 (PS10), 33551 (PS11), 33571 (PS12), 34766 (PS13), 34796 (PS14),

35   34797 (PS15), 34850 (PS16), 35325 (PS17), 35631 (PS18), 35632 (PS19), 35721 (PS20), 35774 (PS21), 35992 (PS22), 36054 (PS23), 36207 (PS24), 36330 (PS25) and 36617 (PS26). The polymorphisms at these sites are guanine or adenine at PS1, cytosine or thymine at PS2, cytosine or

thymine at PS3, cytosine or thymine at PS4, guanine or adenine at PS5, cytosine or thymine at PS6, cytosine or thymine at PS7, adenine or guanine at PS8, cytosine or thymine at PS9, cytosine or guanine at PS10, guanine or adenine at PS11, cytosine or thymine at PS12, cytosine or thymine at PS13, cytosine or thymine at PS14, thymine or cytosine at PS15, cytosine or guanine at PS16, guanine

5   or adenine at PS17, thymine or cytosine at PS18, guanine or thymine at PS19, cytosine or thymine at PS20, thymine or cytosine at PS21, cytosine or thymine at PS22, cytosine or guanine at PS23, cytosine or guanine at PS24, guanine or adenine at PS25 and thymine or cytosine at PS26. In addition, the inventors have determined the identity of the alleles at these sites in a human reference population of 79 unrelated individuals self-identified as belonging to one of four major population groups: African

10   descent, Asian, Caucasian and Hispanic/Latino. From this information, the inventors deduced a set of haplotypes and haplotype pairs for PS1-PS26 in the POR gene, which are shown below in Tables 5 and 4, respectively. Each of these POR haplotypes constitutes a code that defines the variant nucleotides that exist in the human population at this set of polymorphic sites in the POR gene. Thus each POR haplotype also represents a naturally-occurring isoform (also referred to herein as an

15   "isogene") of the POR gene. The frequency of each haplotype and haplotype pair within the total reference population and within each of the four major population groups included in the reference population was also determined.

Thus, in one embodiment, the invention provides a method, composition and kit for genotyping the POR gene in an individual. The genotyping method comprises identifying the

20   nucleotide pair that is present at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18, PS19, PS20, PS21, PS22, PS23, PS24, PS25 and PS26 in both copies of the POR gene from the individual. A genotyping composition of the invention comprises an oligonucleotide probe or primer which is designed to specifically hybridize to a target region containing, or adjacent to, one of these

25   novel POR polymorphic sites. A genotyping kit of the invention comprises a set of oligonucleotides designed to genotype each of these novel POR polymorphic sites. The genotyping method, composition, and kit are useful in determining whether an individual has one of the haplotypes in Table 5 below or has one of the haplotype pairs in Table 4 below.

The invention also provides a method for haplotyping the POR gene in an individual. In one

30   embodiment, the haplotyping method comprises determining, for one copy of the POR gene, the identity of the nucleotide at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18, PS19, PS20, PS21, PS22, PS23, PS24, PS25 and PS26. In another embodiment, the haplotyping method comprises determining whether one copy of the individual's POR gene is defined by one of

35   the POR haplotypes shown in Table 5, below, or a sub-haplotype thereof. In a preferred embodiment, the haplotyping method comprises determining whether both copies of the individual's POR gene are defined by one of the POR haplotype pairs shown in Table 4 below, or a sub-haplotype pair thereof.

Establishing the POR haplotype or haplotype pair of an individual is useful for improving the efficiency and reliability of several steps in the discovery and development of drugs for treating diseases associated with POR activity, e.g., cancers.

For example, the haplotyping method can be used by the pharmaceutical research scientist to

5   validate POR as a candidate target for treating a specific condition or disease predicted to be associated with POR activity. Determining for a particular population the frequency of one or more of the individual POR haplotypes or haplotype pairs described herein will facilitate a decision on whether to pursue POR as a target for treating the specific disease of interest. In particular, if variable POR activity is associated with the disease, then one or more POR haplotypes or haplotype pairs will be

10   found at a higher frequency in disease cohorts than in appropriately genetically matched controls. Conversely, if each of the observed POR haplotypes are of similar frequencies in the disease and control groups, then it may be inferred that variable POR activity has little, if any, involvement with that disease. In either case, the pharmaceutical research scientist can, without *a priori* knowledge as to the phenotypic effect of any POR haplotype or haplotype pair, apply the information derived from

15   detecting POR haplotypes in an individual to decide whether modulating POR activity would be useful in treating the disease.

The claimed invention is also useful in screening for compounds targeting POR to treat a specific condition or disease predicted to be associated with POR activity. For example, detecting which of the POR haplotypes or haplotype-pairs disclosed herein are present in individual members of

20   a population with the specific disease of interest enables the pharmaceutical scientist to screen for a compound(s) that displays the highest desired agonist or antagonist activity for each of the POR isoforms present in the disease population, or for only the most frequent POR isoforms present in the disease population. Thus, without requiring any *a priori* knowledge of the phenotypic effect of any particular POR haplotype or haplotype pair, the claimed haplotyping method provides the scientist

25   with a tool to identify lead compounds that are more likely to show efficacy in clinical trials.

Haplotyping the POR gene in an individual is also useful in the design of clinical trials of candidate drugs for treating a specific condition or disease predicted to be associated with POR activity. For example, instead of randomly assigning patients with the disease of interest to the treatment or control group as is typically done now, determining which of the POR haplotype(s)

30   disclosed herein are present in individual patients enables the pharmaceutical scientist to distribute POR haplotypes and/or haplotype pairs evenly to treatment and control groups, thereby reducing the potential for bias in the results that could be introduced by a larger frequency of a POR haplotype or haplotype pair that is associated with response to the drug being studied in the trial, even if this association was previously unknown. Thus, by practicing the claimed invention, the scientist can

35   more confidently rely on the information learned from the trial, without first determining the phenotypic effect of any POR haplotype or haplotype pair.

In another embodiment, the invention provides a method for identifying an association

between a trait and a POR genotype, haplotype, or haplotype pair for one or more of the novel polymorphic sites described herein. The method comprises comparing the frequency of the POR genotype, haplotype, or haplotype pair in a population exhibiting the trait with the frequency of the POR genotype or haplotype in a reference population. A higher frequency of the POR genotype,

5   haplotype, or haplotype pair in the trait population than in the reference population indicates the trait is associated with the POR genotype, haplotype, or haplotype pair. In preferred embodiments, the trait is susceptibility to a disease, severity of a disease, the staging of a disease or response to a drug. In a particularly preferred embodiment, the POR haplotype is selected from the haplotypes shown in Table 5, or a sub-haplotype thereof. Such methods have applicability in developing diagnostic tests and

10   therapeutic treatments for cancers.

        In yet another embodiment, the invention provides an isolated polynucleotide comprising a nucleotide sequence which is a polymorphic variant of a reference sequence for the POR gene or a fragment thereof. The reference sequence comprises the contiguous sequences shown in Figure 1 and the polymorphic variant comprises at least one polymorphism selected from the group consisting of

15   adenine at PS1, thymine at PS2, thymine at PS3, thymine at PS4, adenine at PS5, thymine at PS6, thymine at PS7, guanine at PS8, thymine at PS9, guanine at PS10, adenine at PS11, thymine at PS12, thymine at PS13, thymine at PS14, cytosine at PS15, guanine at PS16, adenine at PS17, cytosine at PS18, thymine at PS19, thymine at PS20, cytosine at PS21, thymine at PS22, guanine at PS23, guanine at PS24, adenine at PS25 and cytosine at PS26.

20   A particularly preferred polymorphic variant is an isogene of the POR gene. A POR isogene of the invention comprises guanine or adenine at PS1, cytosine or thymine at PS2, cytosine or thymine at PS3, cytosine or thymine at PS4, guanine or adenine at PS5, cytosine or thymine at PS6, cytosine or thymine at PS7, adenine or guanine at PS8, cytosine or thymine at PS9, cytosine or guanine at PS10, guanine or adenine at PS11, cytosine or thymine at PS12, cytosine or thymine at PS13, cytosine or

25   thymine at PS14, thymine or cytosine at PS15, cytosine or guanine at PS16, guanine or adenine at PS17, thymine or cytosine at PS18, guanine or thymine at PS19, cytosine or thymine at PS20, thymine or cytosine at PS21, cytosine or thymine at PS22, cytosine or guanine at PS23, cytosine or guanine at PS24, guanine or adenine at PS25 and thymine or cytosine at PS26. The invention also provides a collection of POR isogenes, referred to herein as a POR genome anthology.

30   In another embodiment, the invention provides a polynucleotide comprising a polymorphic variant of a reference sequence for a POR cDNA or a fragment thereof. The reference sequence comprises SEQ ID NO:2 (Fig.2) and the polymorphic cDNA comprises at least one polymorphism selected from the group consisting of thymine at a position corresponding to nucleotide 266, thymine at a position corresponding to nucleotide 357, thymine at a position corresponding to nucleotide 671,

35   thymine at a position corresponding to nucleotide 675, thymine at a position corresponding to nucleotide 1443, cytosine at a position corresponding to nucleotide 1496 and guanine at a position corresponding to nucleotide 1703. A particularly preferred polymorphic cDNA variant comprises the

coding sequence of a POR isogene defined by haplotypes 1-21.

Polynucleotides complementary to these POR genomic and cDNA variants are also provided by the invention. It is believed that polymorphic variants of the POR gene will be useful in studying the expression and function of POR, and in expressing POR protein for use in screening for candidate drugs to treat diseases related to POR activity.

In other embodiments, the invention provides a recombinant expression vector comprising one of the polymorphic genomic and cDNA variants operably linked to expression regulatory elements as well as a recombinant host cell transformed or transfected with the expression vector. The recombinant vector and host cell may be used to express POR for protein structure analysis and drug binding studies.

In yet another embodiment, the invention provides a polypeptide comprising a polymorphic variant of a reference amino acid sequence for the POR protein. The reference amino acid sequence comprises SEQ ID NO:3 (Fig.3) and the polymorphic variant comprises at least one variant amino acid selected from the group consisting of isoleucine at a position corresponding to amino acid position 89, leucine at a position corresponding to amino acid position 224, alanine at a position corresponding to amino acid position 499 and tryptophan at a position corresponding to amino acid position 568. A polymorphic variant of POR is useful in studying the effect of the variation on the biological activity of POR as well as on the binding affinity of candidate drugs targeting POR for the treatment of cancers.

The present invention also provides antibodies that recognize and bind to the above polymorphic POR protein variant. Such antibodies can be utilized in a variety of diagnostic and prognostic formats and therapeutic methods.

The present invention also provides nonhuman transgenic animals comprising one or more of the POR polymorphic genomic variants described herein and methods for producing such animals. The transgenic animals are useful for studying expression of the POR isogenes *in vivo*, for *in vivo* screening and testing of drugs targeted against POR protein, and for testing the efficacy of therapeutic agents and compounds for cancers in a biological system.

The present invention also provides a computer system for storing and displaying polymorphism data determined for the POR gene. The computer system comprises a computer processing unit; a display; and a database containing the polymorphism data. The polymorphism data includes one or more of the following: the polymorphisms, the genotypes, the haplotypes, and the haplotype pairs identified for the POR gene in a reference population. In a preferred embodiment, the computer system is capable of producing a display showing POR haplotypes organized according to their evolutionary relationships.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 illustrates a reference sequence for the POR gene (Genaissance Reference No.

7

6118814; contiguous lines), with the start and stop positions of each region of coding sequence indicated with a bracket ([or]) and the numerical position below the sequence and the polymorphic site(s) and polymorphism(s) identified by Applicants in a reference population indicated by the variant nucleotide positioned below the polymorphic site in the sequence. SEQ ID NO:1 is equivalent to

5   Figure 1, with the two alternative allelic variants of each polymorphic site indicated by the appropriate nucleotide symbol (R= G or A, Y= T or C, M= A or C, K= G or T, S= G or C, and W= A or T; WIPO standard ST.25). SEQ ID NO:134 is a modified version of SEQ ID NO:1 that shows the context sequence of each polymorphic site, PS1-PS26, in a uniform format to facilitate electronic searching. For each polymorphic site, SEQ ID NO:134 contains a block of 60 bases of the nucleotide sequence

10 encompassing the centrally-located polymorphic site at the 30th position, followed by 60 bases of unspecified sequence to represent that each PS is separated by genomic sequence whose composition is defined elsewhere herein.

Figure 2 illustrates a reference sequence for the POR coding sequence (contiguous lines; SEQ ID NO:2), with the polymorphic site(s) and polymorphism(s) identified by Applicants in a reference

15 population indicated by the variant nucleotide positioned below the polymorphic site in the sequence.

Figure 3 illustrates a reference sequence for the POR protein (contiguous lines; SEQ ID NO:3), with the variant amino acid(s) caused by the polymorphism(s) of Figure 2 positioned below the polymorphic site in the sequence.

20 DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention is based on the discovery of novel variants of the POR gene. As described in more detail below, the inventors herein discovered 21 isogenes of the POR gene by characterizing the POR gene found in genomic DNAs isolated from an Index Repository that contains immortalized cell lines from one chimpanzee and 93 human individuals. The human individuals

25 included a reference population of 79 unrelated individuals self-identified as belonging to one of four major population groups: Caucasian (21 individuals), African descent (20 individuals), Asian (20 individuals), or Hispanic/Latino (18 individuals). To the extent possible, the members of this reference population were organized into population subgroups by their self-identified ethnogeographic origin as shown in Table 1 below.

30

Table 1. Population Groups in the Index Repository

Population Group	Population Subgroup	No. of Individuals
African descent		20
	Sierra Leone	1
Asian		20
	Burma	1
	China	3
	Japan	6
	Korea	1
	Philippines	5
	Vietnam	4
Caucasian		21
	British Isles	3
	British Isles/Central	4
	British Isles/Eastern	1
	Central/Eastern	1
	Eastern	3
	Central/Mediterranean	1
	Mediterranean	2
	Scandinavian	2
Hispanic/Latino		18
	Caribbean	8
	Caribbean (Spanish Descent)	2
	Central American (Spanish Descent)	1
	Mexican American	4
	South American (Spanish Descent)	3

In addition, the Index Repository contains three unrelated indigenous American Indians (one from each of North, Central and South America), one three-generation Caucasian family (from the CEPH Utah cohort) and one two-generation African-American family.

The POR isogenes present in the human reference population are defined by haplotypes for 26 polymorphic sites in the POR gene, all of which are believed to be novel. The novel POR polymorphic sites identified by the inventors are referred to as PS1-PS26 to designate the order in which they are located in the gene (see Table 3 below). Using the genotypes identified in the Index Repository for PS1-PS26 and the methodology described in the Examples below, the inventors herein also determined the pair of haplotypes for the POR gene present in individual human members of this repository. The human genotypes and haplotypes found in the repository for the POR gene include those shown in Tables 4 and 5, respectively. The polymorphism and haplotype data disclosed herein are useful for validating whether POR is a suitable target for drugs to treat cancers, screening for such drugs and reducing bias in clinical trials of such drugs.

In the context of this disclosure, the following terms shall be defined as follows unless otherwise indicated:

Allele - A particular form of a genetic locus, distinguished from other forms by its particular nucleotide sequence.

Candidate Gene – A gene which is hypothesized to be responsible for a disease, condition, or

the response to a treatment, or to be correlated with one of these.

Gene - A segment of DNA that contains all the information for the regulated biosynthesis of an RNA product, including promoters, exons, introns, and other untranslated regions that control expression.

Genotype – An unphased 5´ to 3´ sequence of nucleotide pair(s) found at one or more polymorphic sites in a locus on a pair of homologous chromosomes in an individual. As used herein, genotype includes a full-genotype and/or a sub-genotype as described below.

Full-genotype – The unphased 5´ to 3´ sequence of nucleotide pairs found at all polymorphic sites examined herein in a locus on a pair of homologous chromosomes in a single individual.

Sub-genotype – The unphased 5´ to 3´ sequence of nucleotides seen at a subset of the polymorphic sites examined herein in a locus on a pair of homologous chromosomes in a single individual.

Genotyping – A process for determining a genotype of an individual.

Haplotype – A 5´ to 3´ sequence of nucleotides found at one or more polymorphic sites in a locus on a single chromosome from a single individual. As used herein, haplotype includes a full-haplotype and/or a sub-haplotype as described below.

Full-haplotype – The 5´ to 3´ sequence of nucleotides found at all polymorphic sites examined herein in a locus on a single chromosome from a single individual.

Sub-haplotype – The 5´ to 3´ sequence of nucleotides seen at a subset of the polymorphic sites examined herein in a locus on a single chromosome from a single individual.

Haplotype pair – The two haplotypes found for a locus in a single individual.

Haplotyping – A process for determining one or more haplotypes in an individual and includes use of family pedigrees, molecular techniques and/or statistical inference.

Haplotype data - Information concerning one or more of the following for a specific gene: a listing of the haplotype pairs in each individual in a population; a listing of the different haplotypes in a population; frequency of each haplotype in that or other populations, and any known associations between one or more haplotypes and a trait.

Isoform – A particular form of a gene, mRNA, cDNA or the protein encoded thereby, distinguished from other forms by its particular sequence and/or structure.

Isogene – One of the isoforms (e.g., alleles) of a gene found in a population. An isogene (or allele) contains all of the polymorphisms present in the particular isoform of the gene.

Isolated – As applied to a biological molecule such as RNA, DNA, oligonucleotide, or protein, isolated means the molecule is substantially free of other biological molecules such as nucleic acids, proteins, lipids, carbohydrates, or other material such as cellular debris and growth media. Generally, the term "isolated" is not intended to refer to a complete absence of such material or to absence of water, buffers, or salts, unless they are present in amounts that substantially interfere with the methods of the present invention.

Locus - A location on a chromosome or DNA molecule corresponding to a gene or a physical or phenotypic feature, where physical features include polymorphic sites.

Naturally-occurring – A term used to designate that the object it is applied to, e.g., naturally-occurring polynucleotide or polypeptide, can be isolated from a source in nature and which has not
5   been intentionally modified by man.

Nucleotide pair – The nucleotides found at a polymorphic site on the two copies of a chromosome from an individual.

Phased – As applied to a sequence of nucleotide pairs for two or more polymorphic sites in a locus, phased means the combination of nucleotides present at those polymorphic sites on a single
10   copy of the locus is known.

Polymorphic site (PS) – A position on a chromosome or DNA molecule at which at least two alternative sequences are found in a population.

Polymorphic variant – A gene, mRNA, cDNA, polypeptide or peptide whose nucleotide or amino acid sequence varies from a reference sequence due to the presence of a polymorphism in the
15   gene.

Polymorphism – The sequence variation observed in an individual at a polymorphic site. Polymorphisms include nucleotide substitutions, insertions, deletions and microsatellites and may, but need not, result in detectable differences in gene expression or protein function.

Polymorphism data – Information concerning one or more of the following for a specific
20   gene: location of polymorphic sites; sequence variation at those sites; frequency of polymorphisms in one or more populations; the different genotypes and/or haplotypes determined for the gene; frequency of one or more of these genotypes and/or haplotypes in one or more populations; any known association(s) between a trait and a genotype or a haplotype for the gene.

Polymorphism Database – A collection of polymorphism data arranged in a systematic or
25   methodical way and capable of being individually accessed by electronic or other means.

Polynucleotide – A nucleic acid molecule comprised of single-stranded RNA or DNA or comprised of complementary, double-stranded DNA.

Population Group – A group of individuals sharing a common ethnogeographic origin.

Reference Population – A group of subjects or individuals who are predicted to be
30   representative of the genetic variation found in the general population. Typically, the reference population represents the genetic variation in the population at a certainty level of at least 85%, preferably at least 90%, more preferably at least 95% and even more preferably at least 99%.

Single Nucleotide Polymorphism (SNP) – Typically, the specific pair of nucleotides observed at a single polymorphic site. In rare cases, three or four nucleotides may be found.

35   **Subject** – A human individual whose genotypes or haplotypes or response to treatment or disease state are to be determined.

Treatment - A stimulus administered internally or externally to a subject.

11

Unphased – As applied to a sequence of nucleotide pairs for two or more polymorphic sites in a locus, unphased means the combination of nucleotides present at those polymorphic sites on a single copy of the locus is not known.

5   As discussed above, information on the identity of genotypes and haplotypes for the POR gene of any particular individual as well as the frequency of such genotypes and haplotypes in any particular population of individuals is useful for a variety of drug discovery and development applications. Thus, the invention also provides compositions and methods for detecting the novel POR polymorphisms, haplotypes and haplotype pairs identified herein.

The compositions comprise at least one oligonucleotide for detecting the variant nucleotide or

10   nucleotide pair located at a novel POR polymorphic site in one copy or two copies of the POR gene. Such oligonucleotides are referred to herein as POR haplotyping oligonucleotides or genotyping oligonucleotides, respectively, and collectively as POR oligonucleotides. In one embodiment, a POR haplotyping or genotyping oligonucleotide is a probe or primer capable of hybridizing to a target region that contains, or that is located close to, one of the novel polymorphic sites described herein.

15   As used herein, the term "oligonucleotide" refers to a polynucleotide molecule having less than about 100 nucleotides. A preferred oligonucleotide of the invention is 10 to 35 nucleotides long. More preferably, the oligonucleotide is between 15 and 30, and most preferably, between 20 and 25 nucleotides in length. The exact length of the oligonucleotide will depend on many factors that are routinely considered and practiced by the skilled artisan. The oligonucleotide may be comprised of

20   any phosphorylation state of ribonucleotides, deoxyribonucleotides, and acyclic nucleotide derivatives, and other functionally equivalent derivatives. Alternatively, oligonucleotides may have a phosphate-free backbone, which may be comprised of linkages such as carboxymethyl, acetamidate, carbamate, polyamide (peptide nucleic acid (PNA)) and the like (Varma, R. in Molecular Biology and Biotechnology, A Comprehensive Desk Reference, Ed. R. Meyers, VCH Publishers, Inc. (1995),

25   pages 617-620). Oligonucleotides of the invention may be prepared by chemical synthesis using any suitable methodology known in the art, or may be derived from a biological sample, for example, by restriction digestion. The oligonucleotides may be labeled, according to any technique known in the art, including use of radiolabels, fluorescent labels, enzymatic labels, proteins, haptens, antibodies, sequence tags and the like.

30   Haplotyping or genotyping oligonucleotides of the invention must be capable of specifically hybridizing to a target region of a POR polynucleotide. Preferably, the target region is located in a POR isogene. As used herein, specific hybridization means the oligonucleotide forms an anti-parallel double-stranded structure with the target region under certain hybridizing conditions, while failing to form such a structure when incubated with another region in the POR polynucleotide or with a non-

35   POR polynucleotide under the same hybridizing conditions. Preferably, the oligonucleotide specifically hybridizes to the target region under conventional high stringency conditions. The skilled artisan can readily design and test oligonucleotide probes and primers suitable for detecting

12

polymorphisms in the POR gene using the polymorphism information provided herein in conjunction with the known sequence information for the POR gene and routine techniques.

A nucleic acid molecule such as an oligonucleotide or polynucleotide is said to be a "perfect" or "complete" complement of another nucleic acid molecule if every nucleotide of one of the

5   molecules is complementary to the nucleotide at the corresponding position of the other molecule. A nucleic acid molecule is "substantially complementary" to another molecule if it hybridizes to that molecule with sufficient stability to remain in a duplex form under conventional low-stringency conditions. Conventional hybridization conditions are described, for example, by Sambrook J. et al., in Molecular Cloning, A Laboratory Manual, 2nd Edition, Cold Spring Harbor Press, Cold Spring

10   Harbor, NY (1989) and by Haymes, B.D. et al. in Nucleic Acid Hybridization, A Practical Approach, IRL Press, Washington, D.C. (1985). While perfectly complementary oligonucleotides are preferred for detecting polymorphisms, departures from complete complementarity are contemplated where such departures do not prevent the molecule from specifically hybridizing to the target region. For example, an oligonucleotide primer may have a non-complementary fragment at its 5′ end, with the

15   remainder of the primer being complementary to the target region. Alternatively, non-complementary nucleotides may be interspersed into the probe or primer as long as the resulting probe or primer is still capable of specifically hybridizing to the target region.

Preferred haplotyping or genotyping oligonucleotides of the invention are allele-specific oligonucleotides. As used herein, the term allele-specific oligonucleotide (ASO) means an

20   oligonucleotide that is able, under sufficiently stringent conditions, to hybridize specifically to one allele of a gene, or other locus, at a target region containing a polymorphic site while not hybridizing to the corresponding region in another allele(s). As understood by the skilled artisan, allele-specificity will depend upon a variety of readily optimized stringency conditions, including salt and formamide concentrations, as well as temperatures for both the hybridization and washing steps. Examples of

25   hybridization and washing conditions typically used for ASO probes are found in Kogan et al., "Genetic Prediction of Hemophilia A" in PCR Protocols, A Guide to Methods and Applications, Academic Press, 1990 and Ruaño et al., 87 *Proc. Natl. Acad. Sci. USA* 6296-6300, 1990. Typically, an ASO will be perfectly complementary to one allele while containing a single mismatch for another allele.

30   Allele-specific oligonucleotides of the invention include ASO probes and ASO primers. ASO probes which usually provide good discrimination between different alleles are those in which a central position of the oligonucleotide probe aligns with the polymorphic site in the target region (e.g., approximately the 7th or 8th position in a 15mer, the 8th or 9th position in a 16mer, and the 10th or 11th position in a 20mer). An ASO primer of the invention has a 3′ terminal nucleotide, or preferably a 3′

35   penultimate nucleotide, that is complementary to only one nucleotide of a particular SNP, thereby acting as a primer for polymerase-mediated extension only if the allele containing that nucleotide is present. ASO probes and primers hybridizing to either the coding or noncoding strand are

13

contemplated by the invention. ASO probes and primers listed below use the appropriate nucleotide symbol (R= G or A, Y= T or C, M= A or C, K= G or T, S= G or C, and W= A or T; WIPO standard ST.25) at the position of the polymorphic site to represent that the ASO contains either of the two alternative allelic variants observed at that polymorphic site.

5   A preferred ASO probe for detecting POR gene polymorphisms comprises a nucleotide sequence, listed 5′ to 3′, selected from the group consisting of:

```
        CGTCACCRCATAGAG  (SEQ ID NO:4)  and its complement,
        ACGGGGAYTGCAGAG  (SEQ ID NO:5)  and its complement,
        GGGAGGCYGGCAGGG  (SEQ ID NO:6)  and its complement,
10      TGCAGGCYGACCTGA  (SEQ ID NO:7)  and its complement,
        TCGGGGCRTGCCTGG  (SEQ ID NO:8)  and its complement,
        TTCTGGCYGGCCGTG  (SEQ ID NO:9)  and its complement,
        GGCCGGCYGTGTGTG  (SEQ ID NO:10) and its complement,
        CGGCTCTRTGGCCAC  (SEQ ID NO:11) and its complement,
15      AGTCTGCYGTGTATC  (SEQ ID NO:12) and its complement,
        CCAGAAGSGTCCTTG  (SEQ ID NO:13) and its complement,
        TGGAGACRGAGACTC  (SEQ ID NO:14) and its complement,
        AAAGCCCYGGCCGCT  (SEQ ID NO:15) and its complement,
        CTGTTGCYGCAGAGC  (SEQ ID NO:16) and its complement,
20      CCCCCTCYTGCCGCA  (SEQ ID NO:17) and its complement,
        CCCCTCCYGCCGCAG  (SEQ ID NO:18) and its complement,
        TGCCTGTSTCTTCCC  (SEQ ID NO:19) and its complement,
        GCCACACRCTGGAGG  (SEQ ID NO:20) and its complement,
        GCCTCGGYGTGGCGG  (SEQ ID NO:21) and its complement,
25      CCTCGGTKTGGCGGT  (SEQ ID NO:22) and its complement,
        CCAAGGCYGGCCGCA  (SEQ ID NO:23) and its complement,
        GAGCCTGYCGGGGAG  (SEQ ID NO:24) and its complement,
        CTCACCCYGGCCCCT  (SEQ ID NO:25) and its complement,
        CGCCGCTSGGATGAG  (SEQ ID NO:26) and its complement,
30      GCCACAGSCACAGTG  (SEQ ID NO:27) and its complement,
        TGTGGGTRAGTGAGT  (SEQ ID NO:28) and its complement, and
        TCAGCTCYCCTGGCT  (SEQ ID NO:29) and its complement.
```

A preferred ASO primer for detecting POR gene polymorphisms comprises a nucleotide 35   sequence, listed 5′ to 3′, selected from the group consisting of:

```
        GGCTGGCGTCACCRC  (SEQ ID NO:30);  GACATCCTCTATGYG  (SEQ ID NO:31);
40      TCCCAGACGGGGAYT  (SEQ ID NO:32);  AAACTCCTCTGCART  (SEQ ID NO:33);
        AGCAGGGGGAGGCYG  (SEQ ID NO:34);  CCCACTCCCTGCCRG  (SEQ ID NO:35);
        TGTTTCTGCAGGCYG  (SEQ ID NO:36);  GGCTGCTCAGGTCRG  (SEQ ID NO:37);
        GTGGGGTCGGGGCRT  (SEQ ID NO:38);  CTGGTGCCAGGCAYG  (SEQ ID NO:39);
        GAGCAGTTCTGGCYG  (SEQ ID NO:40);  TTCACACACGGCCRG  (SEQ ID NO:41);
45      AGTTCTGGCCGGCYG  (SEQ ID NO:42);  AGTGTTCACACACRG  (SEQ ID NO:43);
        GGCAGACGGCTCTRT  (SEQ ID NO:44);  GCACCAGTGGCCAYA  (SEQ ID NO:45);
        AGGCTCAGTCTGCYG  (SEQ ID NO:46);  TATGGGGATACACRG  (SEQ ID NO:47);
        GTGCAACCAGAAGSG  (SEQ ID NO:48);  CGTCTCCAAGGACSC  (SEQ ID NO:49);
        CGTCCTTGGAGACRG  (SEQ ID NO:50);  TGATCTGAGTCTCYG  (SEQ ID NO:51);
50      CAGATCAAAGCCCYG  (SEQ ID NO:52);  ACAGTGAGCGGCCRG  (SEQ ID NO:53);
        GGGCACCTGTTGCYG  (SEQ ID NO:54);  GGGCCAGCTCTGCRG  (SEQ ID NO:55);
        GTGTCACCCCCTCYT  (SEQ ID NO:56);  GGTGGCTGCGGCARG  (SEQ ID NO:57);
```

```
     TGTCACCCCCTCCYG (SEQ ID NO:58); GGGTGGCTGCGGCRG (SEQ ID NO:59);
     AAGTCCTGCCTGTST (SEQ ID NO:60); TCTGCAGGGAAGASA (SEQ ID NO:61);
     CTGCCAGCCACACRC (SEQ ID NO:62); GCTGGGCCTCCAGYG (SEQ ID NO:63);
     GCAAGGGCCTCGGYG (SEQ ID NO:64); GCTCCACCGCCACRC (SEQ ID NO:65);
5    CAAGGGCCTCGGTKT (SEQ ID NO:66); AGCTCCACCGCCAMA (SEQ ID NO:67);
     ACGAGACCAAGGCYG (SEQ ID NO:68); TGTTGATGCGGCCRG (SEQ ID NO:69);
     GCCAAGGAGCCTGYC (SEQ ID NO:70); GCCGTTCTCCCCGRC (SEQ ID NO:71);
     CCCGCGCTCACCCYG (SEQ ID NO:72); CGTGGCAGGGGCCRG (SEQ ID NO:73);
     GGCTGCCGCCGCTSG (SEQ ID NO:74); GTAGTCCTCATCCSA (SEQ ID NO:75);
10   GGCAGGGCCACAGSC (SEQ ID NO:76); GGGGGGCACTGTGSC (SEQ ID NO:77);
     TACGTCTGTGGGTRA (SEQ ID NO:78); GACCCCACTCACTYA (SEQ ID NO:79);
     CTGTAATCAGCTCYC (SEQ ID NO:80); and GGAGGGAGCCAGGRG (SEQ ID NO:81).
```

Other oligonucleotides of the invention hybridize to a target region located one to several

15 nucleotides downstream of one of the novel polymorphic sites identified herein. Such

oligonucleotides are useful in polymerase-mediated primer extension methods for detecting one of the

novel polymorphisms described herein and therefore such oligonucleotides are referred to herein as

"primer-extension oligonucleotides". In a preferred embodiment, the 3'-terminus of a primer-

extension oligonucleotide is a deoxynucleotide complementary to the nucleotide located immediately

20 adjacent to the polymorphic site.

A particularly preferred oligonucleotide primer for detecting POR gene polymorphisms by

primer extension terminates in a nucleotide sequence, listed 5' to 3', selected from the group

consisting of:

```
     TGGCGTCACC  (SEQ ID NO:82);  ATCCTCTATG  (SEQ ID NO:83);
25   CAGACGGGGA  (SEQ ID NO:84);  CTCCTCTGCA  (SEQ ID NO:85);
     AGGGGGAGGC  (SEQ ID NO:86);  ACTCCCTGCC  (SEQ ID NO:87);
     TTCTGCAGGC  (SEQ ID NO:88);  TGCTCAGGTC  (SEQ ID NO:89);
     GGGTCGGGGC  (SEQ ID NO:90);  GTGCCAGGCA  (SEQ ID NO:91);
     CAGTTCTGGC  (SEQ ID NO:92);  ACACACGGCC  (SEQ ID NO:93);
30   TCTGGCCGGC  (SEQ ID NO:94);  GTTCACACAC  (SEQ ID NO:95);
     AGACGGCTCT  (SEQ ID NO:96);  CCAGTGGCCA  (SEQ ID NO:97);
     CTCAGTCTGC  (SEQ ID NO:98);  GGGGATACAC  (SEQ ID NO:99);
     CAACCAGAAG  (SEQ ID NO:100); CTCCAAGGAC  (SEQ ID NO:101);
     CCTTGGAGAC  (SEQ ID NO:102); TCTGAGTCTC  (SEQ ID NO:103);
35   ATCAAAGCCC  (SEQ ID NO:104); GTGAGCGGCC  (SEQ ID NO:105);
     CACCTGTTGC  (SEQ ID NO:106); CCAGCTCTGC  (SEQ ID NO:107);
     TCACCCCCTC  (SEQ ID NO:108); GGCTGCGGCA  (SEQ ID NO:109);
     CACCCCCTCC  (SEQ ID NO:110); TGGCTGCGGC  (SEQ ID NO:111);
     TCCTGCCTGT  (SEQ ID NO:112); GCAGGGAAGA  (SEQ ID NO:113);
40   CCAGCCACAC  (SEQ ID NO:114); GGGCCTCCAG  (SEQ ID NO:115);
     AGGGCCTCGG  (SEQ ID NO:116); CCACCGCCAC  (SEQ ID NO:117);
     GGGCCTCGGT  (SEQ ID NO:118); TCCACCGCCA  (SEQ ID NO:119);
     AGACCAAGGC  (SEQ ID NO:120); TGATGCGGCC  (SEQ ID NO:121);
     AAGGAGCCTG  (SEQ ID NO:122); GTTCTCCCCG  (SEQ ID NO:123);
45   GCGCTCACCC  (SEQ ID NO:124); GGCAGGGGCC  (SEQ ID NO:125);
     TGCCGCCGCT  (SEQ ID NO:126); GTCCTCATCC  (SEQ ID NO:127);
     AGGGCCACAG  (SEQ ID NO:128); GGGCACTGTG  (SEQ ID NO:129);
     GTCTGTGGGT  (SEQ ID NO:130); CCCACTCACT  (SEQ ID NO:131);
     TAATCAGCTC  (SEQ ID NO:132); and GGGAGCCAGG  (SEQ ID NO:133).
50
```

In some embodiments, a composition contains two or more differently labeled POR

oligonucleotides for simultaneously probing the identity of nucleotides or nucleotide pairs at two or more polymorphic sites. It is also contemplated that primer compositions may contain two or more sets of allele-specific primer pairs to allow simultaneous targeting and amplification of two or more regions containing a polymorphic site.

5   POR oligonucleotides of the invention may also be immobilized on or synthesized on a solid surface such as a microchip, bead, or glass slide (see, e.g., WO 98/20020 and WO 98/20019). Such immobilized oligonucleotides may be used in a variety of polymorphism detection assays, including but not limited to probe hybridization and polymerase extension assays. Immobilized POR oligonucleotides of the invention may comprise an ordered array of oligonucleotides designed to

10   rapidly screen a DNA sample for polymorphisms in multiple genes at the same time.

      In another embodiment, the invention provides a kit comprising at least two POR oligonucleotides packaged in separate containers. The kit may also contain other components such as hybridization buffer (where the oligonucleotides are to be used as a probe) packaged in a separate container. Alternatively, where the oligonucleotides are to be used to amplify a target region, the kit

15   may contain, packaged in separate containers, a polymerase and a reaction buffer optimized for primer extension mediated by the polymerase, such as PCR.

      The above described oligonucleotide compositions and kits are useful in methods for genotyping and/or haplotyping the POR gene in an individual. As used herein, the terms "POR genotype" and "POR haplotype" mean the genotype or haplotype contains the nucleotide pair or

20   nucleotide, respectively, that is present at one or more of the novel polymorphic sites described herein and may optionally also include the nucleotide pair or nucleotide present at one or more additional polymorphic sites in the POR gene. The additional polymorphic sites may be currently known polymorphic sites or sites that are subsequently discovered.

      One embodiment of a genotyping method of the invention involves isolating from the

25   individual a nucleic acid sample comprising the two copies of the POR gene, mRNA transcripts thereof or cDNA copies thereof, or a fragment of any of the foregoing, that are present in the individual, and determining the identity of the nucleotide pair at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18, PS19, PS20, PS21, PS22, PS23, PS24, PS25 and PS26 in

30   the two copies to assign a POR genotype to the individual. As will be readily understood by the skilled artisan, the two "copies" of a gene, mRNA or cDNA (or fragment of such POR molecules) in an individual may be the same allele or may be different alleles. In another embodiment, a genotyping method of the invention comprises determining the identity of the nucleotide pair at each of PS1-PS26.

35   Typically, the nucleic acid sample is isolated from a biological sample taken from the individual, such as a blood sample or tissue sample. Suitable tissue samples include whole blood, semen, saliva, tears, urine, fecal material, sweat, buccal, skin and hair. The nucleic acid sample may

be comprised of genomic DNA, mRNA, or cDNA and, in the latter two cases, the biological sample must be obtained from a tissue in which the POR gene is expressed. Furthermore it will be understood by the skilled artisan that mRNA or cDNA preparations would not be used to detect polymorphisms located in introns or in 5′ and 3′ untranslated regions if not present in the mRNA or cDNA. If a POR

5   gene fragment is isolated, it must contain the polymorphic site(s) to be genotyped.

One embodiment of a haplotyping method of the invention comprises isolating from the individual a nucleic acid sample containing only one of the two copies of the POR gene, mRNA or cDNA, or a fragment of such POR molecules, that is present in the individual and determining in that copy the identity of the nucleotide at one or more polymorphic sites selected from the group consisting

10   of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18, PS19, PS20, PS21, PS22, PS23, PS24, PS25 and PS26 in that copy to assign a POR haplotype to the individual.

The nucleic acid used in the above haplotyping methods of the invention may be isolated using any method capable of separating the two copies of the POR gene or fragment such as one of the

15   methods described above for preparing POR isogenes, with targeted *in vivo* cloning being the preferred approach. As will be readily appreciated by those skilled in the art, any individual clone will typically only provide haplotype information on one of the two POR gene copies present in an individual. If haplotype information is desired for the individual's other copy, additional POR clones will usually need to be examined. Typically, at least five clones should be examined to have more

20   than a 90% probability of haplotyping both copies of the POR gene in an individual. In some cases, however, once the haplotype for one POR allele is directly determined, the haplotype for the other allele may be inferred if the individual has a known genotype for the polymorphic sites of interest or if the haplotype frequency or haplotype pair frequency for the individual's population group is known. In a particularly preferred embodiment, the nucleotide at each of PS1-PS26 is identified.

25   In another embodiment, the haplotyping method comprises determining whether an individual has one or more of the POR haplotypes shown in Table 5. This can be accomplished by identifying, for one or both copies of the individual's POR gene, the phased sequence of nucleotides present at each of PS1-PS26. This identifying step does not necessarily require that each of PS1-PS26 be directly examined. Typically only a subset of PS1-PS26 will need to be directly examined to assign to

30   an individual one or more of the haplotypes shown in Table 5. This is because at least one polymorphic site in a gene is frequently in strong linkage disequilibrium with one or more other polymorphic sites in that gene (Drysdale, CM et al. 2000 *PNAS* 97:10483-10488; Rieder MJ et al. 1999 *Nature Genetics* 22:59-62). Two sites are said to be in linkage disequilibrium if the presence of a particular variant at one site enhances the predictability of another variant at the second site

35   (Stephens, JC 1999, *Mol. Diag.* 4:309-317). Techniques for determining whether any two polymorphic sites are in linkage disequilibrium are well-known in the art (Weir B.S. 1996 *Genetic Data Analysis II*, Sinauer Associates, Inc. Publishers, Sunderland, MA).

In another embodiment of a haplotyping method of the invention, a POR haplotype pair is determined for an individual by identifying the phased sequence of nucleotides at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18, PS19, PS20, PS21, PS22, PS23, PS24,

5   PS25 and PS26 in each copy of the POR gene that is present in the individual. In a particularly preferred embodiment, the haplotyping method comprises identifying the phased sequence of nucleotides at each of PS1-PS26 in each copy of the POR gene.

When haplotyping both copies of the gene, the identifying step is preferably performed with each copy of the gene being placed in separate containers. However, it is also envisioned that if the

10   two copies are labeled with different tags, or are otherwise separately distinguishable or identifiable, it could be possible in some cases to perform the method in the same container. For example, if first and second copies of the gene are labeled with different first and second fluorescent dyes, respectively, and an allele-specific oligonucleotide labeled with yet a third different fluorescent dye is used to assay the polymorphic site(s), then detecting a combination of the first and third dyes would identify the

15   polymorphism in the first gene copy while detecting a combination of the second and third dyes would identify the polymorphism in the second gene copy.

In both the genotyping and haplotyping methods, the identity of a nucleotide (or nucleotide pair) at a polymorphic site(s) may be determined by amplifying a target region(s) containing the polymorphic site(s) directly from one or both copies of the POR gene, or a fragment thereof, and the

20   sequence of the amplified region(s) determined by conventional methods. It will be readily appreciated by the skilled artisan that only one nucleotide will be detected at a polymorphic site in individuals who are homozygous at that site, while two different nucleotides will be detected if the individual is heterozygous for that site. The polymorphism may be identified directly, known as positive-type identification, or by inference, referred to as negative-type identification. For example,

25   where a SNP is known to be guanine and cytosine in a reference population, a site may be positively determined to be either guanine or cytosine for an individual homozygous at that site, or both guanine and cytosine, if the individual is heterozygous at that site. Alternatively, the site may be negatively determined to be not guanine (and thus cytosine/cytosine) or not cytosine (and thus guanine/guanine).

The target region(s) may be amplified using any oligonucleotide-directed amplification

30   method, including but not limited to polymerase chain reaction (PCR) (U.S. Patent No. 4,965,188), ligase chain reaction (LCR) (Barany et al., *Proc. Natl. Acad. Sci. USA* 88:189-193, 1991; WO90/01069), and oligonucleotide ligation assay (OLA) (Landegren et al., *Science* 241:1077-1080, 1988). Other known nucleic acid amplification procedures may be used to amplify the target region including transcription-based amplification systems (U.S. Patent No. 5,130,238; EP 329,822; U.S.

35   Patent No. 5,169,766, WO89/06700) and isothermal methods (Walker et al., *Proc. Natl. Acad. Sci. USA* 89:392-396, 1992).

A polymorphism in the target region may also be assayed before or after amplification using

one of several hybridization-based methods known in the art. Typically, allele-specific oligonucleotides are utilized in performing such methods. The allele-specific oligonucleotides may be used as differently labeled probe pairs, with one member of the pair showing a perfect match to one variant of a target sequence and the other member showing a perfect match to a different variant. In

5   some embodiments, more than one polymorphic site may be detected at once using a set of allele-specific oligonucleotides or oligonucleotide pairs. Preferably, the members of the set have melting temperatures within 5°C, and more preferably within 2°C, of each other when hybridizing to each of the polymorphic sites being detected.

Hybridization of an allele-specific oligonucleotide to a target polynucleotide may be

10   performed with both entities in solution, or such hybridization may be performed when either the oligonucleotide or the target polynucleotide is covalently or noncovalently affixed to a solid support. Attachment may be mediated, for example, by antibody-antigen interactions, poly-L-Lys, streptavidin or avidin-biotin, salt bridges, hydrophobic interactions, chemical linkages, UV cross-linking baking, etc. Allele-specific oligonucleotides may be synthesized directly on the solid support or attached to

15   the solid support subsequent to synthesis. Solid-supports suitable for use in detection methods of the invention include substrates made of silicon, glass, plastic, paper and the like, which may be formed, for example, into wells (as in 96-well plates), slides, sheets, membranes, fibers, chips, dishes, and beads. The solid support may be treated, coated or derivatized to facilitate the immobilization of the allele-specific oligonucleotide or target nucleic acid.

20   The genotype or haplotype for the POR gene of an individual may also be determined by hybridization of a nucleic acid sample containing one or both copies of the gene, mRNA, cDNA or fragment(s) thereof, to nucleic acid arrays and subarrays such as described in WO 95/11995. The arrays would contain a battery of allele-specific oligonucleotides representing each of the polymorphic sites to be included in the genotype or haplotype.

25   The identity of polymorphisms may also be determined using a mismatch detection technique, including but not limited to the RNase protection method using riboprobes (Winter et al., *Proc. Natl. Acad. Sci. USA* 82:7575, 1985; Meyers et al., *Science* 230:1242, 1985) and proteins which recognize nucleotide mismatches, such as the E. coli mutS protein (Modrich, P. *Ann. Rev. Genet.* 25:229-253, 1991). Alternatively, variant alleles can be identified by single strand conformation polymorphism

30   (SSCP) analysis (Orita et al., *Genomics* 5:874-879, 1989; Humphries et al., in Molecular Diagnosis of Genetic Diseases, R. Elles, ed., pp. 321-340, 1996) or denaturing gradient gel electrophoresis (DGGE) (Wartell et al., *Nucl. Acids Res.* 18:2699-2706, 1990; Sheffield et al., *Proc. Natl. Acad. Sci. USA* 86:232-236, 1989).

A polymerase-mediated primer extension method may also be used to identify the

35   polymorphism(s). Several such methods have been described in the patent and scientific literature and include the "Genetic Bit Analysis" method (WO92/15712) and the ligase/polymerase mediated genetic bit analysis (U.S. Patent 5,679,524. Related methods are disclosed in WO91/02087, WO90/09455,

WO95/17676, U.S. Patent Nos. 5,302,509, and 5,945,283. Extended primers containing a polymorphism may be detected by mass spectrometry as described in U.S. Patent No. 5,605,798. Another primer extension method is allele-specific PCR (Ruaño et al., *Nucl. Acids Res.* 17:8392, 1989; Ruaño et al., *Nucl. Acids Res.* 19, 6877-6882, 1991; WO 93/22456; Turki et al., *J. Clin. Invest.*

5   95:1635-1641, 1995). In addition, multiple polymorphic sites may be investigated by simultaneously amplifying multiple regions of the nucleic acid using sets of allele-specific primers as described in Wallace et al. (WO89/10414).

In addition, the identity of the allele(s) present at any of the novel polymorphic sites described herein may be indirectly determined by haplotyping or genotyping another polymorphic site that is in

10   linkage disequilibrium with the polymorphic site that is of interest. Polymorphic sites in linkage disequilibrium with the presently disclosed polymorphic sites may be located in regions of the gene or in other genomic regions not examined herein. Detection of the allele(s) present at a polymorphic site in linkage disequilibrium with the novel polymorphic sites described herein may be performed by, but is not limited to, any of the above-mentioned methods for detecting the identity of the allele at a

15   polymorphic site.

In another aspect of the invention, an individual's POR haplotype pair is predicted from its POR genotype using information on haplotype pairs known to exist in a reference population. In its broadest embodiment, the haplotyping prediction method comprises identifying a POR genotype for the individual at two or more POR polymorphic sites described herein, accessing data containing POR

20   haplotype pairs identified in a reference population, and assigning a haplotype pair to the individual that is consistent with the genotype data. In one embodiment, the reference haplotype pairs include the POR haplotype pairs shown in Table 4. The POR haplotype pair can be assigned by comparing the individual's genotype with the genotypes corresponding to the haplotype pairs known to exist in the general population or in a specific population group, and determining which haplotype pair is

25   consistent with the genotype of the individual. In some embodiments, comparison of the genotype of the individual to the haplotype pairs identified in a reference population and determination of which haplotype pair is consistent with the genotype of the individual may be performed by visual inspection (for example, by consulting Table 4). When the genotype of the individual is consistent with more than one haplotype pair, haplotype pair frequency data (such as that presented in Table 7) may be used

30   to determine which of these haplotype pairs is most likely to be present in the individual. This determination may also be performed in some embodiments by visual inspection upon consulting Table 7. If a particular POR haplotype pair consistent with the genotype of the individual is more frequent in the reference population than others consistent with the genotype, then that haplotype pair with the highest frequency is the most likely to be present in the individual. In other embodiments, the

35   comparison may be made by a computer-implemented algorithm with the genotype of the individual and the reference haplotype data stored in computer-readable formats. For example, as described in PCT/US01/12831, filed April 18, 2001, one computer-implemented algorithm to perform this

comparison entails enumerating all possible haplotype pairs which are consistent with the genotype, accessing data containing POR haplotype pairs frequency data determined in a reference population to determine a probability that the individual has a possible haplotype pair, and analyzing the determined probabilities to assign a haplotype pair to the individual.

5   Generally, the reference population should be composed of randomly-selected individuals representing the major ethnogeographic groups of the world. A preferred reference population for use in the methods of the present invention comprises an approximately equal number of individuals from Caucasian, African-descent, Asian and Hispanic-Latino population groups with the minimum number of each group being chosen based on how rare a haplotype one wants to be guaranteed to see. For

10   example, if one wants to have a q% chance of not missing a haplotype that exists in the population at a p% frequency of occurring in the reference population, the number of individuals (n) who must be sampled is given by 2n=log(1-q)/log(1-p) where p and q are expressed as fractions. A preferred reference population allows the detection of any haplotype whose frequency is at least 10% with about 99% certainty and comprises about 20 unrelated individuals from each of the four population groups

15   named above. A particularly preferred reference population includes a 3-generation family representing one or more of the four population groups to serve as controls for checking quality of haplotyping procedures.

        In a preferred embodiment, the haplotype frequency data for each ethnogeographic group is examined to determine whether it is consistent with Hardy-Weinberg equilibrium. Hardy-Weinberg

20   equilibrium (D.L. Hartl et al., Principles of Population Genomics, Sinauer Associates (Sunderland, MA), 3rd Ed., 1997) postulates that the frequency of finding the haplotype pair H_1 / H_2 is equal to

$$p_{H-W}(H_1 / H_2) = 2p(H_1)p(H_2) \text{ if } H_1 \neq H_2 \text{ and } p_{H-W}(H_1 / H_2) = p(H_1)p(H_2) \text{ if } H_1 = H_2.$$

A statistically significant difference between the observed and expected haplotype frequencies could be due to one or more factors including significant inbreeding in the population group, strong selective

25   pressure on the gene, sampling bias, and/or errors in the genotyping process. If large deviations from Hardy-Weinberg equilibrium are observed in an ethnogeographic group, the number of individuals in that group can be increased to see if the deviation is due to a sampling bias. If a larger sample size does not reduce the difference between observed and expected haplotype pair frequencies, then one may wish to consider haplotyping the individual using a direct haplotyping method such as, for

30   example, CLASPER System™ technology (U.S. Patent No. 5,866,404), single molecule dilution, or allele-specific long-range PCR (Michalotos-Beloin et al., *Nucleic Acids Res.* 24:4841-4843, 1996).

        In one embodiment of this method for predicting a POR haplotype pair for an individual, the assigning step involves performing the following analysis. First, each of the possible haplotype pairs is compared to the haplotype pairs in the reference population. Generally, only one of the haplotype

35   pairs in the reference population matches a possible haplotype pair and that pair is assigned to the individual. Occasionally, only one haplotype represented in the reference haplotype pairs is consistent with a possible haplotype pair for an individual, and in such cases the individual is assigned a

21

haplotype pair containing this known haplotype and a new haplotype derived by subtracting the known haplotype from the possible haplotype pair. Alternatively, the haplotype pair in an individual may be predicted from the individual's genotype for that gene using reported methods (e.g., Clark et al. 1990 *Mol Bio Evol* 7:111-22; copending PCT/US01/12831 filed April 18, 2001) or through a

5   commercial haplotyping service such as offered by Genaissance Pharmaceuticals, Inc. (New Haven, CT). In rare cases, either no haplotypes in the reference population are consistent with the possible haplotype pairs, or alternatively, multiple reference haplotype pairs are consistent with the possible haplotype pairs. In such cases, the individual is preferably haplotyped using a direct molecular haplotyping method such as, for example, CLASPER System™ technology (U.S. Patent No.

10   5,866,404), SMD, or allele-specific long-range PCR (Michalotos-Beloin et al., *supra*).

The invention also provides a method for determining the frequency of a POR genotype, haplotype, or haplotype pair in a population. The method comprises, for each member of the population, determining the genotype or the haplotype pair for the novel POR polymorphic sites described herein, and calculating the frequency any particular genotype; haplotype, or haplotype pair

15   is found in the population. The population may be e.g., a reference population, a family population, a same gender population, a population group, or a trait population (e.g., a group of individuals exhibiting a trait of interest such as a medical condition or response to a therapeutic treatment).

In another aspect of the invention, frequency data for POR genotypes, haplotypes, and/or haplotype pairs are determined in a reference population and used in a method for identifying an

20   association between a trait and a POR genotype, haplotype, or haplotype pair. The trait may be any detectable phenotype, including but not limited to susceptibility to a disease or response to a treatment. In one embodiment, the method involves obtaining data on the frequency of the genotype(s), haplotype(s), or haplotype pair(s) of interest in a reference population as well as in a population exhibiting the trait. Frequency data for one or both of the reference and trait populations may be

25   obtained by genotyping or haplotyping each individual in the populations using one or more of the methods described above. The haplotypes for the trait population may be determined directly or, alternatively, by a predictive genotype to haplotype approach as described above. In another embodiment, the frequency data for the reference and/or trait populations is obtained by accessing previously determined frequency data, which may be in written or electronic form. For example, the

30   frequency data may be present in a database that is accessible by a computer. Once the frequency data is obtained, the frequencies of the genotype(s), haplotype(s), or haplotype pair(s) of interest in the reference and trait populations are compared. In a preferred embodiment, the frequencies of all genotypes, haplotypes, and/or haplotype pairs observed in the populations are compared. If a particular POR genotype, haplotype, or haplotype pair is more frequent in the trait population than in

35   the reference population at a statistically significant amount, then the trait is predicted to be associated with that POR genotype, haplotype or haplotype pair. Preferably, the POR genotype, haplotype, or haplotype pair being compared in the trait and reference populations is selected from the full-

genotypes and full-haplotypes shown in Tables 4 and 5, or from sub-genotypes and sub-haplotypes derived from these genotypes and haplotypes.

In a preferred embodiment of the method, the trait of interest is a clinical response exhibited by a patient to some therapeutic treatment, for example, response to a drug targeting POR or response

5   to a therapeutic treatment for a medical condition. As used herein, "medical condition" includes but is not limited to any condition or disease manifested as one or more physical and/or psychological symptoms for which treatment is desirable, and includes previously and newly identified diseases and other disorders. As used herein the term "clinical response" means any or all of the following: a quantitative measure of the response, no response, and/or adverse response (i.e., side effects).

10   In order to deduce a correlation between clinical response to a treatment and a POR genotype, haplotype, or haplotype pair, it is necessary to obtain data on the clinical responses exhibited by a population of individuals who received the treatment, hereinafter the "clinical population". This clinical data may be obtained by analyzing the results of a clinical trial that has already been run and/or the clinical data may be obtained by designing and carrying out one or more new clinical trials.

15   As used herein, the term "clinical trial" means any research study designed to collect clinical data on. responses to a particular treatment, and includes but is not limited to phase I, phase II and phase III clinical trials. Standard methods are used to define the patient population and to enroll subjects.

It is preferred that the individuals included in the clinical population have been graded for the existence of the medical condition of interest. This is important in cases where the symptom(s) being

20   presented by the patients can be caused by more than one underlying condition, and where treatment of the underlying conditions are not the same. An example of this would be where patients experience breathing difficulties that are due to either asthma or respiratory infections. If both sets were treated with an asthma medication, there would be a spurious group of apparent non-responders that did not actually have asthma. These people would affect the ability to detect any correlation between

25   haplotype and treatment outcome. This grading of potential patients could employ a standard physical exam or one or more lab tests. Alternatively, grading of patients could use haplotyping for situations where there is a strong correlation between haplotype pair and disease susceptibility or severity.

The therapeutic treatment of interest is administered to each individual in the trial population and each individual's response to the treatment is measured using one or more predetermined criteria.

30   It is contemplated that in many cases, the trial population will exhibit a range of responses and that the investigator will choose the number of responder groups (e.g., low, medium, high) made up by the various responses. In addition, the POR gene for each individual in the trial population is genotyped and/or haplotyped, which may be done before or after administering the treatment.

After both the clinical and polymorphism data have been obtained, correlations between

35   individual response and POR genotype or haplotype content are created. Correlations may be produced in several ways. In one method, individuals are grouped by their POR genotype or haplotype (or haplotype pair) (also referred to as a polymorphism group), and then the averages and

standard deviations of clinical responses exhibited by the members of each polymorphism group are calculated.

These results are then analyzed to determine if any observed variation in clinical response between polymorphism groups is statistically significant. Statistical analysis methods which may be
5 used are described in L.D. Fisher and G. vanBelle, "Biostatistics: A Methodology for the Health Sciences", Wiley-Interscience (New York) 1993. This analysis may also include a regression calculation of which polymorphic sites in the POR gene give the most significant contribution to the differences in phenotype. One regression model useful in the invention is described in WO 01/01218, entitled "Methods for Obtaining and Using Haplotype Data".

10 A second method for finding correlations between POR haplotype content and clinical responses uses predictive models based on error-minimizing optimization algorithms. One of many possible optimization algorithms is a genetic algorithm (R. Judson, "Genetic Algorithms and Their Uses in Chemistry" in Reviews in Computational Chemistry, Vol. 10, pp. 1-73, K. B. Lipkowitz and D. B. Boyd, eds. (VCH Publishers, New York, 1997). Simulated annealing (Press et al., "Numerical
15 Recipes in C: The Art of Scientific Computing", Cambridge University Press (Cambridge) 1992, Ch. 10), neural networks (E. Rich and K. Knight, "Artificial Intelligence", 2nd Edition (McGraw-Hill, New York, 1991, Ch. 18), standard gradient descent methods (Press et al., *supra*, Ch. 10), or other global or local optimization approaches (see discussion in Judson, *supra*) could also be used. Preferably, the correlation is found using a genetic algorithm approach as described in WO 01/01218.

20 Correlations may also be analyzed using analysis of variation (ANOVA) techniques to determine how much of the variation in the clinical data is explained by different subsets of the polymorphic sites in the POR gene. As described in WO 01/01218, ANOVA is used to test hypotheses about whether a response variable is caused by or correlated with one or more traits or variables that can be measured (Fisher and vanBelle, *supra*, Ch. 10).

25 From the analyses described above, a mathematical model may be readily constructed by the skilled artisan that predicts clinical response as a function of POR genotype or haplotype content. Preferably, the model is validated in one or more follow-up clinical trials designed to test the model.

The identification of an association between a clinical response and a genotype or haplotype (or haplotype pair) for the POR gene may be the basis for designing a diagnostic method to determine
30 those individuals who will or will not respond to the treatment, or alternatively, will respond at a lower level and thus may require more treatment, i.e., a greater dose of a drug. The diagnostic method may take one of several forms: for example, a direct DNA test (i.e., genotyping or haplotyping one or more of the polymorphic sites in the POR gene), a serological test, or a physical exam measurement. The only requirement is that there be a good correlation between the diagnostic test results and the
35 underlying POR genotype or haplotype that is in turn correlated with the clinical response. In a preferred embodiment, this diagnostic method uses the predictive haplotyping method described above.

In another embodiment, the invention provides an isolated polynucleotide comprising a polymorphic variant of the POR gene or a fragment of the gene which contains at least one of the novel polymorphic sites described herein. The nucleotide sequence of a variant POR gene is identical to the reference genomic sequence for those portions of the gene examined, as described in the

5   Examples below, except that it comprises a different nucleotide at one or more of the novel polymorphic sites PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18, PS19, PS20, PS21, PS22, PS23, PS24, PS25 and PS26. Similarly, the nucleotide sequence of a variant fragment of the POR gene is identical to the corresponding portion of the reference sequence except for having a different nucleotide at one or more of the novel

10   polymorphic sites described herein. Thus, the invention specifically does not include polynucleotides comprising a nucleotide sequence identical to the reference sequence of the POR gene (or other reported POR sequences) or to portions of the reference sequence (or other reported POR sequences), except for the haplotyping and genotyping oligonucleotides described above.

The location of a polymorphism in a variant POR gene or fragment is preferably identified by

15   aligning its sequence against SEQ ID NO:1. The polymorphism is selected from the group consisting of adenine at PS1, thymine at PS2, thymine at PS3, thymine at PS4, adenine at PS5, thymine at PS6, thymine at PS7, guanine at PS8, thymine at PS9, guanine at PS10, adenine at PS11, thymine at PS12, thymine at PS13, thymine at PS14, cytosine at PS15, guanine at PS16, adenine at PS17, cytosine at PS18, thymine at PS19, thymine at PS20, cytosine at PS21, thymine at PS22, guanine at PS23,

20   guanine at PS24, adenine at PS25 and cytosine at PS26. In a preferred embodiment, the polymorphic variant comprises a naturally-occurring isogene of the POR gene which is defined by any one of haplotypes 1-21 shown in Table 5 below.

Polymorphic variants of the invention may be prepared by isolating a clone containing the POR gene from a human genomic library. The clone may be sequenced to determine the identity of

25   the nucleotides at the novel polymorphic sites described herein. Any particular variant or fragment thereof, that is claimed herein could be prepared from this clone by performing *in vitro* mutagenesis using procedures well-known in the art. Any particular POR variant or fragment thereof may also be prepared using synthetic or semi-synthetic methods known in the art.

POR isogenes, or fragments thereof, may be isolated using any method that allows separation

30   of the two "copies" of the POR gene present in an individual, which, as readily understood by the skilled artisan, may be the same allele or different alleles. Separation methods include targeted *in vivo* cloning (TIVC) in yeast as described in WO 98/01573, U.S. Patent No. 5,866,404, and U.S. Patent No. 5,972,614. Another method, which is described in U.S. Patent No. 5,972,614, uses an allele specific oligonucleotide in combination with primer extension and exonuclease degradation to generate

35   hemizygous DNA targets. Yet other methods are single molecule dilution (SMD) as described in Ruaño et al., *Proc. Natl. Acad. Sci.* 87:6296-6300, 1990; and allele specific PCR (Ruaño et al., 1989, *supra*; Ruaño et al., 1991, *supra*; Michalatos-Beloin et al., *supra*).

The invention also provides POR genome anthologies, which are collections of at least two POR isogenes found in a given population. The population may be any group of at least two individuals, including but not limited to a reference population, a population group, a family population, a clinical population, and a same gender population. A POR genome anthology may

5   comprise individual POR isogenes stored in separate containers such as microtest tubes, separate wells of a microtitre plate and the like. Alternatively, two or more groups of the POR isogenes in the anthology may be stored in separate containers. Individual isogenes or groups of such isogenes in a genome anthology may be stored in any convenient and stable form, including but not limited to in buffered solutions, as DNA precipitates, freeze-dried preparations and the like. A preferred POR

10   genome anthology of the invention comprises a set of isogenes defined by the haplotypes shown in Table 5 below. A POR genome anthology is useful in providing control nucleic acids for kits of the invention.

An isolated polynucleotide containing a polymorphic variant nucleotide sequence of the invention may be operably linked to one or more expression regulatory elements in a recombinant

15   expression vector capable of being propagated and expressing the encoded POR protein in a prokaryotic or a eukaryotic host cell. Examples of expression regulatory elements which may be used include, but are not limited to, the lac system, operator and promoter regions of phage lambda, yeast promoters, and promoters derived from vaccinia virus, adenovirus, retroviruses, or SV40. Other regulatory elements include, but are not limited to, appropriate leader sequences, termination codons,

20   polyadenylation signals, and other sequences required for the appropriate transcription and subsequent translation of the nucleic acid sequence in a given host cell. Of course, the correct combinations of expression regulatory elements will depend on the host system used. In addition, it is understood that the expression vector contains any additional elements necessary for its transfer to and subsequent replication in the host cell. Examples of such elements include, but are not limited to, origins of

25   replication and selectable markers. Such expression vectors are commercially available or are readily constructed using methods known to those in the art (e.g., F. Ausubel et al., 1987, in "Current Protocols in Molecular Biology", John Wiley and Sons, New York, New York). Host cells which may be used to express the variant POR sequences of the invention include, but are not limited to, eukaryotic and mammalian cells, such as animal, plant, insect and yeast cells, and prokaryotic cells,

30   such as E. coli, or algal cells as known in the art. The recombinant expression vector may be · introduced into the host cell using any method known to those in the art including, but not limited to, microinjection, electroporation, particle bombardment, transduction, and transfection using DEAE-dextran, lipofection, or calcium phosphate (see e.g., Sambrook et al. (1989) in "Molecular Cloning. A Laboratory Manual", Cold Spring Harbor Press, Plainview, New York). In a preferred aspect,

35   eukaryotic expression vectors that function in eukaryotic cells, and preferably mammalian cells, are used. Non-limiting examples of such vectors include vaccinia virus vectors, adenovirus vectors, herpes virus vectors, and baculovirus transfer vectors. Preferred eukaryotic cell lines include COS

cells, CHO cells, HeLa cells, NIH/3T3 cells, and embryonic stem cells (Thomson, J. A. et al., 1998 *Science* 282:1145-1147). Particularly preferred host cells are mammalian cells.

As will be readily recognized by the skilled artisan, expression of polymorphic variants of the POR gene will produce POR mRNAs varying from each other at any polymorphic site retained in the

5   spliced and processed mRNA molecules. These mRNAs can be used for the preparation of a POR cDNA comprising a nucleotide sequence which is a polymorphic variant of the POR reference coding sequence shown in Figure 2. Thus, the invention also provides POR mRNAs and corresponding cDNAs which comprise a nucleotide sequence that is identical to SEQ ID NO:2 (Fig. 2), or its corresponding RNA sequence, for those regions of SEQ ID NO:2 that correspond to the examined

10   portions of the POR gene (as described in the Examples below), except for having one or more polymorphisms selected from the group consisting of thymine at a position corresponding to nucleotide 266, thymine at a position corresponding to nucleotide 357, thymine at a position corresponding to nucleotide 671, thymine at a position corresponding to nucleotide 675, thymine at a position corresponding to nucleotide 1443, cytosine at a position corresponding to nucleotide 1496

15   and guanine at a position corresponding to nucleotide 1703. A particularly preferred polymorphic cDNA variant comprises the coding sequence of a POR isogene defined by any one of haplotypes 1-21. Fragments of these variant mRNAs and cDNAs are included in the scope of the invention, provided they contain one or more of the novel polymorphisms described herein. The invention specifically excludes polynucleotides identical to previously identified and characterized POR

20   mRNAs, cDNAs or fragments thereof. Polynucleotides comprising a variant POR RNA or DNA sequence may be isolated from a biological sample using well-known molecular biological procedures or may be chemically synthesized.

As used herein, a polymorphic variant of a POR gene, mRNA or cDNA fragment comprises at least one novel polymorphism identified herein and has a length of at least 10 nucleotides and may

25   range up to the full length of the gene. Preferably, such fragments are between 100 and 3000 nucleotides in length, and more preferably between 200 and 2000 nucleotides in length, and most preferably between 500 and 1000 nucleotides in length.

In describing the POR polymorphic sites identified herein, reference is made to the sense strand of the gene for convenience. However, as recognized by the skilled artisan, nucleic acid

30   molecules containing the POR gene or cDNA may be complementary double stranded molecules and thus reference to a particular site on the sense strand refers as well to the corresponding site on the complementary antisense strand. Thus, reference may be made to the same polymorphic site on either strand and an oligonucleotide may be designed to hybridize specifically to either strand at a target region containing the polymorphic site. Thus, the invention also includes single-stranded

35   polynucleotides which are complementary to the sense strand of the POR genomic, mRNA and cDNA variants described herein.

Polynucleotides comprising a polymorphic gene variant or fragment of the invention may be

27

useful for therapeutic purposes. For example, where a patient could benefit from expression, or increased expression, of a particular POR protein isoform, an expression vector encoding the isoform may be administered to the patient. The patient may be one who lacks the POR isogene encoding that isoform or may already have at least one copy of that isogene.

5   - In other situations, it may be desirable to decrease or block expression of a particular POR isogene. Expression of a POR isogene may be turned off by transforming a targeted organ, tissue or cell population with an expression vector that expresses high levels of untranslatable mRNA or antisense RNA for the isogene or fragment thereof. Alternatively, oligonucleotides directed against the regulatory regions (e.g., promoter, introns, enhancers, 3′ untranslated region) of the isogene may

10   block transcription. Oligonucleotides targeting the transcription initiation site, e.g., between positions −10 and +10 from the start site are preferred. Similarly, inhibition of transcription can be achieved using oligonucleotides that base-pair with region(s) of the isogene DNA to form triplex DNA (see e.g., Gee et al. in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco, N.Y., 1994). Antisense oligonucleotides may also be designed to block translation of

15   POR mRNA transcribed from a particular isogene. It is also contemplated that ribozymes may be designed that can catalyze the specific cleavage of POR mRNA transcribed from a particular isogene.

          The untranslated mRNA, antisense RNA or antisense oligonucleotides may be delivered to a target cell or tissue by expression from a vector introduced into the cell or tissue *in vivo* or *ex vivo*. Alternatively, such molecules may be formulated as a pharmaceutical composition for administration

20   to the patient. Oligoribonucleotides and/or oligodeoxynucleotides intended for use as antisense oligonucleotides may be modified to increase stability and half-life. Possible modifications include, but are not limited to phosphorothioate or 2′ O-methyl linkages, and the inclusion of nontraditional bases such as inosine and queosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytosine, guanine, thymine, and uracil which are not as easily recognized by endogenous

25   nucleases.

          The invention also provides an isolated polypeptide comprising a polymorphic variant of (a) the reference POR amino acid sequence shown in Figure 3 or (b) a fragment of this reference sequence. The location of a variant amino acid in a POR polypeptide or fragment of the invention is identified by aligning its sequence against SEQ ID NO:3 (Fig. 3). A POR protein variant of the

30   invention comprises an amino acid sequence identical to SEQ ID NO:3 for those regions of SEQ ID NO:3 that are encoded by examined portions of the POR gene (as described in the Examples below), except for having one or more variant amino acids selected from the group consisting of isoleucine at a position corresponding to amino acid position 89, leucine at a position corresponding to amino acid position 224, alanine at a position corresponding to amino acid position 499 and tryptophan at a

35   position corresponding to amino acid position 568. Thus, a POR fragment of the invention, also referred to herein as a POR peptide variant, is any fragment of a POR protein variant that contains one or more of the amino acid variations shown in Table 2. The invention specifically excludes amino

28

acid sequences identical to those previously identified for POR, including SEQ ID NO:3, and previously described fragments thereof. POR protein variants included within the invention comprise all amino acid sequences based on SEQ ID NO:3 and having the combination of amino acid variations described in Table 2 below. In preferred embodiments, a POR protein variant of the invention is

5 encoded by an isogene defined by one of the observed haplotypes, 1-21, shown in Table 5.

Table 2. Novel Polymorphic Variants of POR

Polymorphic Variant Number	Amino Acid Position and Identities			
	89	224	499	568
1	T	P	V	W
2	T	P	A	S
3	T	P	A	W
4	T	L	V	S
5	T	L	V	W.
6	T	L	A	S
7	T	L	A	W
8	I	P	V	S
9	I	P	V	W
10	I	P	A	S
11	I	P	A	W
12	I	L	V	S
13	I	L	V	W
14	I	L	A	S
15	I	L	A	W

A POR peptide variant of the invention is at least 6 amino acids in length and is preferably any number between 6 and 30 amino acids long, more preferably between 10 and 25, and most

30 preferably between 15 and 20 amino acids long. Such POR peptide variants may be useful as antigens to generate antibodies specific for one of the above POR isoforms. In addition, the POR peptide variants may be useful in drug screening assays.

A POR variant protein or peptide of the invention may be prepared by chemical synthesis or by expressing an appropriate variant POR genomic or cDNA sequence described above.

35 Alternatively, the POR protein variant may be isolated from a biological sample of an individual having a POR isogene which encodes the variant protein. Where the sample contains two different POR isoforms (i.e., the individual has different POR isogenes), a particular POR isoform of the invention can be isolated by immunoaffinity chromatography using an antibody which specifically binds to that particular POR isoform but does not bind to the other POR isoform.

40 The expressed or isolated POR protein or peptide may be detected by methods known in the art, including Coomassie blue staining, silver staining, and Western blot analysis using antibodies specific for the isoform of the POR protein or peptide as discussed further below. POR variant proteins and peptides can be purified by standard protein purification procedures known in the art, including differential precipitation, molecular sieve chromatography, ion-exchange chromatography,

isoelectric focusing, gel electrophoresis, affinity and immunoaffinity chromatography and the like. (Ausubel et. al., 1987, In Current Protocols in Molecular Biology John Wiley and Sons, New York, New York). In the case of immunoaffinity chromatography, antibodies specific for a particular polymorphic variant may be used.

5   A polymorphic variant POR gene of the invention may also be fused in frame with a heterologous sequence to encode a chimeric POR protein. The non-POR portion of the chimeric protein may be recognized by a commercially available antibody. In addition, the chimeric protein may also be engineered to contain a cleavage site located between the POR and non-POR portions so that the POR protein may be cleaved and purified away from the non-POR portion.

10   An additional embodiment of the invention relates to using a novel POR protein isoform, or a fragment thereof, in any of a variety of drug screening assays. Such screening assays may be performed to identify agents that bind specifically to all known POR protein isoforms or to only a subset of one or more of these isoforms. The agents may be from chemical compound libraries, peptide libraries and the like. The POR protein or peptide variant may be free in solution or affixed to

15   a solid support. In one embodiment, high throughput screening of compounds for binding to a POR variant may be accomplished using the method described in PCT application WO84/03565, in which large numbers of test compounds are synthesized on a solid substrate, such as plastic pins or some other surface, contacted with the POR protein(s) of interest and then washed. Bound POR protein(s) are then detected using methods well-known in the art.

20   In another embodiment, a novel POR protein isoform may be used in assays to measure the binding affinities of one or more candidate drugs targeting the POR protein.

            In yet another embodiment, when a particular POR haplotype or group of POR haplotypes encodes a POR protein variant with an amino acid sequence distinct from that of POR protein isoforms encoded by other POR haplotypes, then detection of that particular POR haplotype or group

25   of POR haplotypes may be accomplished by detecting expression of the encoded POR protein variant using any of the methods described herein or otherwise commonly known to the skilled artisan.

            In another embodiment, the invention provides antibodies specific for and immunoreactive with one or more of the novel POR variant proteins described herein. The antibodies may be either monoclonal or polyclonal in origin. The POR protein or peptide variant used to generate the antibodies

30   may be from natural or recombinant sources or produced by chemical synthesis using synthesis techniques known in the art. If the POR protein variant is of insufficient size to be antigenic, it may be conjugated, complexed, or otherwise covalently linked to a carrier molecule to enhance the antigenicity of the peptide. Examples of carrier molecules, include, but are not limited to, albumins (e.g., human, bovine, fish, ovine), and keyhole limpet hemocyanin (Basic and Clinical Immunology,

35   1991, Eds. D.P. Stites, and A.I. Terr, Appleton and Lange, Norwalk Connecticut, San Mateo, California).

            In one embodiment, an antibody specifically immunoreactive with one of the novel protein

isoforms described herein is administered to an individual to neutralize activity of the POR isoform expressed by that individual. The antibody may be formulated as a pharmaceutical composition which includes a pharmaceutically acceptable carrier.

Antibodies specific for and immunoreactive with one of the novel protein isoforms described
5   herein may be used to immunoprecipitate the POR protein variant from solution as well as react with POR protein isoforms on Western or immunoblots of polyacrylamide gels on membrane supports or substrates. In another preferred embodiment, the antibodies will detect POR protein isoforms in paraffin or frozen tissue sections, or in cells which have been fixed or unfixed and prepared on slides, coverslips, or the like, for use in immunocytochemical, immunohistochemical, and
10 immunofluorescence techniques.

In another embodiment, an antibody specifically immunoreactive with one of the novel POR protein variants described herein is used in immunoassays to detect this variant in biological samples. In this method, an antibody of the present invention is contacted with a biological sample and the formation of a complex between the POR protein variant and the antibody is detected. As described,
15 suitable immunoassays include radioimmunoassay, Western blot assay, immunofluorescent assay, enzyme linked immunoassay (ELISA), chemiluminescent assay, immunohistochemical assay, immunocytochemical assay, and the like (see, e.g., Principles and Practice of Immunoassay, 1991, Eds. Christopher P. Price and David J. Neoman, Stockton Press, New York, New York; Current Protocols in Molecular Biology, 1987, Eds. Ausubel et al., John Wiley and Sons, New York, New
20 York). Standard techniques known in the art for ELISA are described in Methods in Immunodiagnosis, 2nd Ed., Eds. Rose and Bigazzi, John Wiley and Sons, New York 1980; and Campbell et al., 1984, Methods in Immunology, W.A. Benjamin, Inc.). Such assays may be direct, indirect, competitive, or noncompetitive as described in the art (see, e.g., Principles and Practice of Immunoassay, 1991, Eds. Christopher P. Price and David J. Neoman, Stockton Pres, NY, NY; and
25 Oellirich, M., 1984, J. Clin. Chem. Clin. Biochem., 22:895-904). Proteins may be isolated from test specimens and biological samples by conventional methods, as described in Current Protocols in Molecular Biology, supra.

Exemplary antibody molecules for use in the detection and therapy methods of the present invention are intact immunoglobulin molecules, substantially intact immunoglobulin molecules, or
30 those portions of immunoglobulin molecules that contain the antigen binding site. Polyclonal or monoclonal antibodies may be produced by methods conventionally known in the art (e.g., Kohler and Milstein, 1975, Nature, 256:495-497; Campbell Monoclonal Antibody Technology, the Production and Characterization of Rodent and Human Hybridomas, 1985, In: Laboratory Techniques in Biochemistry and Molecular Biology, Eds. Burdon et al., Volume 13, Elsevier Science Publishers,
35 Amsterdam). The antibodies or antigen binding fragments thereof may also be produced by genetic engineering. The technology for expression of both heavy and light chain genes in E. coli is the subject of PCT patent applications, publication number WO 901443, WO 901443 and WO 9014424

and in Huse et al., 1989, Science, 246:1275-1281. The antibodies may also be humanized (e.g., Queen, C. et al. 1989 Proc. Natl. Acad. Sci.USA 86;10029).

Effect(s) of the polymorphisms identified herein on expression of POR may be investigated by preparing recombinant cells and/or nonhuman recombinant organisms, preferably recombinant
5   animals, containing a polymorphic variant of the POR gene. As used herein, "expression" includes but is not limited to one or more of the following: transcription of the gene into precursor mRNA; splicing and other processing of the precursor mRNA to produce mature mRNA; mRNA stability; translation of the mature mRNA into POR protein (including codon usage and tRNA availability); and glycosylation and/or other modifications of the translation product, if required for proper expression
10 and function.

To prepare a recombinant cell of the invention, the desired POR isogene may be introduced into the cell in a vector such that the isogene remains extrachromosomal. In such a situation, the gene will be expressed by the cell from the extrachromosomal location. In a preferred embodiment, the POR isogene is introduced into a cell in such a way that it recombines with the endogenous POR gene
15 present in the cell. Such recombination requires the occurrence of a double recombination event, thereby resulting in the desired POR gene polymorphism. Vectors for the introduction of genes both for recombination and for extrachromosomal maintenance are known in the art, and any suitable vector or vector construct may be used in the invention. Methods such as electroporation, particle bombardment, calcium phosphate co-precipitation and viral transduction for introducing DNA into
20 cells are known in the art; therefore, the choice of method may lie with the competence and preference of the skilled practitioner. Examples of cells into which the POR isogene may be introduced include, but are not limited to, continuous culture cells, such as COS, NIH/3T3, and primary or culture cells of the relevant tissue type, i.e., they express the POR isogene. Such recombinant cells can be used to compare the biological activities of the different protein variants.
25 Recombinant nonhuman organisms, i.e., transgenic animals, expressing a variant POR gene are prepared using standard procedures known in the art. Preferably, a construct comprising the variant gene is introduced into a nonhuman animal or an ancestor of the animal at an embryonic stage, i.e., the one-cell stage, or generally not later than about the eight-cell stage. Transgenic animals carrying the constructs of the invention can be made by several methods known to those having skill
30 in the art. One method involves transfecting into the embryo a retrovirus constructed to contain one or more insulator elements, a gene or genes of interest, and other components known to those skilled in the art to provide a complete shuttle vector harboring the insulated gene(s) as a transgene, see e.g., U.S. Patent No. 5,610,053. Another method involves directly injecting a transgene into the embryo. A third method involves the use of embryonic stem cells. Examples of animals into which the POR
35 isogenes may be introduced include, but are not limited to, mice, rats, other rodents, and nonhuman primates (see "The Introduction of Foreign Genes into Mice" and the cited references therein, In: Recombinant DNA, Eds. J.D. Watson, M. Gilman, J. Witkowski, and M. Zoller; W.H. Freeman and

32

Company, New York, pages 254-272). Transgenic animals stably expressing a human POR isogene and producing the encoded human POR protein can be used as biological models for studying diseases related to abnormal POR expression and/or activity, and for screening and assaying various candidate drugs, compounds, and treatment regimens to reduce the symptoms or effects of these diseases.

5   An additional embodiment of the invention relates to pharmaceutical compositions for treating disorders affected by expression or function of a novel POR isogene described herein. The pharmaceutical composition may comprise any of the following active ingredients: a polynucleotide comprising one of these novel POR isogenes; an antisense oligonucleotide directed against one of the novel POR isogenes, a polynucleotide encoding such an antisense oligonucleotide, or another

10   compound which inhibits expression of a novel POR isogene described herein. Preferably, the composition contains the active ingredient in a therapeutically effective amount. By therapeutically effective amount is meant that one or more of the symptoms relating to disorders affected by expression or function of a novel POR isogene is reduced and/or eliminated. The composition also comprises a pharmaceutically acceptable carrier, examples of which include, but are not limited to,

15   saline, buffered saline, dextrose, and water. Those skilled in the art may employ a formulation most suitable for the active ingredient, whether it is a polynucleotide, oligonucleotide, protein, peptide or small molecule antagonist. The pharmaceutical composition may be administered alone or in combination with at least one other agent, such as a stabilizing compound. Administration of the pharmaceutical composition may be by any number of routes including, but not limited to oral,

20   intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, intradermal, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing Co., Easton, PA).

For any composition, determination of the therapeutically effective dose of active ingredient

25   and/or the appropriate route of administration is well within the capability of those skilled in the art. For example, the dose can be estimated initially either in cell culture assays or in animal models. The animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans. The exact dosage will be determined by the practitioner, in light of factors

30   relating to the patient requiring treatment, including but not limited to severity of the disease state, general health, age, weight and gender of the patient, diet, time and frequency of administration, other drugs being taken by the patient, and tolerance/response to the treatment.

Any or all analytical and mathematical operations involved in practicing the methods of the present invention may be implemented by a computer. In addition, the computer may execute a

35   program that generates views (or screens) displayed on a display device and with which the user can interact to view and analyze large amounts of information relating to the POR gene and its genomic variation, including chromosome location, gene structure, and gene family, gene expression data,

polymorphism data, genetic sequence data, and clinical data population data (e.g., data on ethnogeographic origin, clinical responses, genotypes, and haplotypes for one or more populations). The POR polymorphism data described herein may be stored as part of a relational database (e.g., an instance of an Oracle database or a set of ASCII flat files). These polymorphism data may be stored

5   on the computer's hard drive or may, for example, be stored on a CD-ROM or on one or more other storage devices accessible by the computer. For example, the data may be stored on one or more databases in communication with the computer via a network.

Preferred embodiments of the invention are described in the following examples. Other embodiments within the scope of the claims herein will be apparent to one skilled in the art from

10   consideration of the specification or practice of the invention as disclosed herein. It is intended that the specification, together with the examples, be considered exemplary only, with the scope and spirit of the invention being indicated by the claims which follow the examples.

EXAMPLES

15   The Examples herein are meant to exemplify the various aspects of carrying out the invention and are not intended to limit the scope of the invention in any way. The Examples do not include detailed descriptions for conventional methods employed, such as in the performance of genomic DNA isolation, PCR and sequencing procedures. Such methods are well-known to those skilled in the art and are described in numerous publications, for example, Sambrook, Fritsch, and Maniatis,

20   "Molecular Cloning: A Laboratory Manual", 2nd Edition, Cold Spring Harbor Laboratory Press, USA, (1989).

EXAMPLE 1

This example illustrates examination of various regions of the POR gene for polymorphic

25   sites.

Amplification of Target Regions

The following target regions of the POR gene were amplified using PCR primer pairs. The primers used for each region are represented below by providing the nucleotide positions of their

30   initial and final nucleotides, which correspond to positions in SEQ ID NO:1 (Figure 1).

PCR Primer Pairs

Fragment No.	Forward Primer	Reverse Primer	PCR Product
Fragment 1	3878-3899	complement of 4429-4406	552 nt
Fragment 2	22306-22328	complement of 22748-22726	443 nt
Fragment 3	29363-29385	complement of 29820-29801	458 nt
Fragment 4	30215-30236	complement of 30747-30725	533 nt
Fragment 5	30942-30962	complement of 31433-31413	492 nt
Fragment 6	31414-31434	complement of 31856-31836	443 nt
Fragment 7	32039-32059	complement of 32600-32579	562 nt
Fragment 8	33448-33468	complement of 33880-33859	433 nt
Fragment 9	33639-33658	complement of 34167-34145	529 nt
Fragment 10	34661-34683	complement of 35288-35265	628 nt
Fragment 11	34915-34939	complement of 35578-35559	664 nt
Fragment 12	35497-35516	complement of 36043-36022	547 nt
Fragment 13	35664-35683	complement of 36162-36142	499 nt
Fragment 14	35847-35866	complement of 36348-36327	502 nt
Fragment 15	36051-36074	complement of 36479-36458	429 nt
Fragment 16	36117-36136	complement of 36857-36835	741 nt

These primer pairs were used in PCR reactions containing genomic DNA isolated from immortalized cell lines for each member of the Index Repository. The PCR reactions were carried out under the following conditions:

Reaction volume	= 10 µl
10 x Advantage 2 Polymerase reaction buffer (Clontech)	= 1 µl
100 ng of human genomic DNA	= 1 µl
10 mM dNTP	= 0.4 µl
Advantage 2 Polymerase enzyme mix (Clontech)	= 0.2 µl
Forward Primer (10 µM)	= 0.4 µl
Reverse Primer (10 µM)	= 0.4 µl
Water	= 6.6 µl

Amplification profile:

97°C - 2 min.   1 cycle

97°C - 15 sec.
70°C - 45 sec.   }   10 cycles
72°C - 45 sec.

97°C - 15 sec.
64°C - 45 sec.   }   35 cycles
72°C - 45 sec.

Sequencing of PCR Products

The PCR products were purified using a Whatman/Polyfiltronics 100 µl 384 well unifilter plate essentially according to the manufacturers protocol. The purified DNA was eluted in 50 µl of distilled water. Sequencing reactions were set up using Applied Biosystems Big Dye Terminator chemistry essentially according to the manufacturers protocol. The purified PCR products were

sequenced in both directions using the primer sets represented below by the nucleotide positions of

their initial and final nucleotides, which correspond to positions in SEQ ID NO:1 (Figure 1). Reaction

products were purified by isopropanol precipitation, and run on an Applied Biosystems 3700 DNA

Analyzer.

5

Sequencing Primer Pairs

Fragment No.	Forward Primer	Reverse Primer
Fragment 1	3944-3963	complement of 4392-4372
Fragment 2	22375-22394	complement of 22722-22703
Fragment 3	29406-29426	complement of 29787-29768
Fragment 4	30292-30312	complement of 30725-30706
Fragment 5	30968-30987	complement of 31389-31370
Fragment 6	31442-31461	complement of 31822-31803
Fragment 7	32174-32193	complement of 32547-32527
Fragment 8	33480-33499	complement of 33845-33826
Fragment 9	33681-33700	complement of 34099-34080
Fragment 10	34730-34749	complement of 35169-35150
Fragment 11	35004-35021	complement of 35443-35425
Fragment 12	35534-35553	complement of 35915-35895
Fragment 13	35692-35711	complement of 36081-36060
Fragment 14	35885-35904	complement of 36290-36271
Fragment 15	36117-36136	complement of 36449-36431
Fragment 16	36300-36318	complement of 36804-36790

10

15

20

25   ## Analysis of Sequences for Polymorphic Sites

Sequence information for a minimum of 80 humans was analyzed for the presence of

polymorphisms using the Polyphred program (Nickerson et al., *Nucleic Acids Res.* 14:2745-2751,

1997). The presence of a polymorphism was confirmed on both strands. The polymorphisms and their

locations in the POR reference genomic sequence (SEQ ID NO:1) are listed in Table 3 below.

Table 3. Polymorphic Sites Identified in the POR Gene

Polymorphic Site Number	PolyId(a)	Nucleotide Position	Reference Allele	Variant Allele	CDS Variant Position	AA Variant
PS1	6664331	22635	G	A		
PS2	6664339	29577	C	T	266	T89I
PS3	6664341	29754	C	T		
PS4	6664343	30427	C	T	357	A119A
PS5	6664355	31531	G	A		
PS6	6664359	31644	C	T	671	P224L
PS7	6664361	31648	C	T	675	A225A
PS8	6664371	32446	A	G		
PS9	6664373	32478	C	T		
PS10	6664377	33538	C	G		
PS11	6664379	33551	G	A		
PS12	6664381	33571	C	T		
PS13	6664393	34766	C	T		
PS14	6664395	34796	C	T		
PS15	6664397	34797	T	C		
PS16	6664399	34850	C	G		
PS17	6664409	35325	G	A		
PS18	6664413	35631	T	C		
PS19	6664415	35632	G	T		
PS20	6664417	35721	C	T	1443	A481A
PS21	6664419	35774	T	C	1496	V499A
PS22	6664425	35992	C	T		
PS23	6664429	36054	C	G	1703	S568W
PS24	6664433	36207	C	G		
PS25	6664441	36330	G	A		
PS26	6664443	36617	T	C		

(a) PolyId is a unique identifier assigned to each PS by Genaissance Pharmaceuticals, Inc.

EXAMPLE 2

This example illustrates analysis of the POR polymorphisms identified in the Index Repository for human genotypes and haplotypes.

The different genotypes containing these polymorphisms that were observed in unrelated members of the reference population are shown in Table 4 below, with the haplotype pair indicating the combination of haplotypes determined for the individual using the haplotype derivation protocol described below. In Table 4, homozygous positions are indicated by one nucleotide and heterozygous positions are indicated by two nucleotides. Missing nucleotides in any given genotype in Table 4 were inferred based on linkage disequilibrium and/or Mendelian inheritance.

Table 4 (Part 1). Genotypes and Haplotype Pairs Observed for POR Gene

Genotype Number	HAP Pair		PS1	PS2	PS3	PS4	PS5	PS6	PS7	PS8	PS9	PS10
1	8	8	G	C	C	C	A	C	C	A	C	G
2	10	10	G	C	C	C	A	C	C	A	C	G
3	19	19	G	C	T	C	A	C	C	A	C	G
4	15	15	G	C	C	C	G	C	C	A	C	C
5	8	11	G	C	C	C	A	C	C	A	C	G
6	8	3	G/A	C	C	C	A	C	C	A	C	G
7	19	21	G	C/T	T/C	C	A	C	C	A	C	G
8	10	17	G	C	C	C	A/G	C	C	A	C	G/C
9	15	2	G/A	C	C	C	G/A	C	C	A	C	C/G
10	19	14	G	C	T/C	C	A/G	C	C	A	C	G/C
11	8	16	G	C	C	C	A/G	C	C	A	C	G/C
12	8	20	G	C	C/T	C	A	C	C	A	C/T	G
13	19	2	G/A	C	T/C	C	A	C	C	A	C	G
14	19	4	G/A	C	T	C	A	C	C	A	C	G
15	14	9	G	C	C	C	G/A	C	C	A	C	C/G
16	8	13	G	C	C	C	A	C	C/T	A	C	G
17	8	12	G	C	C	C	A	C	C	A/G	C	G
18	8	1	G/A	C	C	C	A	C	C	A	C	G
19	8	6	G	C	C	C	A	C	C	A	C	G
20	8	15	G	C	C	C	A/G	C	C	A	C	G/C
21	8	17	G	C	C	C	A/G	C	C	A	C	G/C
22	8	7	G	C	C	C	A	C	C	A	C	G
23	10	5	G	C	C	C	A	C	C	A	C	G
24	8	2	G/A	C	C	C	A	C	C	A	C	G
25	8	19	G	C	C/T	C	A	C	C	A	C	G
26	10	15	G	C	C	C	A/G	C	C	A	C	G/C
27	8	10	G	C	C	C	A	C	C	A	C	G
28	10	19	G	C	C/T	C	A	C	C	A	C	G
29	19	17	G	C	T/C	C	A/G	C	C	A	C	G/C
30	15	18	G	C	C	C/T	G	C	C	A	C	C
31	19	15	G	C	T/C	C	A/G	C	C	A	C	G/C

Table 4 (Part 2). Genotypes and Haplotype Pairs Observed for POR Gene

Genotype Number	HAP Pair		PS11	PS12	PS13	PS14	PS15	PS16	PS17	PS18	PS19	PS20
1	8	8	A	C	T	C	C	G	G	C	T	C
2	10	10	A	T	C	C	C	G	G	C	T	C
3	19	19	A	C	T	C	C	G	G	C	T	C
4	15	15	G	C	C	C	T	C	G	T	G	T
5	8	11	A	C/T	T	C	C	G	G	C	T	C
6	8	3	A	C/T	T/C	C	C	G	G/A	C/T	T	C
7	19	21	A	C	T	C	C	G	G	C	T	C
8	10	17	A/G	T/C	C	C/T	C/T	G/C	G	C/T	T/G	C/T
9	15	2	G/A	C/T	C	C	T/C	C/G	G/A	T/C	G/T	T/C
10	19	14	A/G	C	T/C	C	C/T	G/C	G	C/T	T/G	C
11	8	16	A/G	C	T/C	C	C/T	G	G	C/T	T/G	C/T
12	8	20	A	C	T	C	C	G	G	C	T	C
13	19	2	A	C/T	T/C	C	C	G	G/A	C	T	C
14	19	4	A	C	T	C	C	G	G	C	T	C
15	14	9	G/A	C/T	C	C	T/C	C/G	G	T/C	G/T	C
16	8	13	A	C/T	T/C	C	C	G	G	C	T	C
17	8	12	A	C/T	T/C	C	C	G	G	C	T	C
18	8	1	A	C	T	C	C	G	G	C	T	C
19	8	6	A	C	T/C	C	C	G	G	C	T	C
20	8	15	A/G	C	T/C	C	C/T	G/C	G	C/T	T/G	C/T
21	8	17	A/G	C	T/C	C/T	C/T	G/C	G	C/T	T/G	C/T
22	8	7	A	C	T	C	C	G	G	C	T	C
23	10	5	A	T/C	C	C	C	G	G	C	T	C
24	8	2	A	C/T	T/C	C	C	G	G/A	C	T	C
25	8	19	A	C	T	C	C	G	G	C	T	C
26	10	15	A/G	T/C	C	C	C/T	G/C	G	C/T	T/G	C/T
27	8	10	A	C/T	T/C	C	C	G	G	C	T	C
28	10	19	A	T/C	C/T	C	C	G	G	C	T	C
29	19	17	A/G	C	T/C	C/T	C/T	G/C	G	C/T	T/G	C/T
30	15	18	G	C	C	C	T	C	G	T	G	T/C
31	19	15	A/G	C	T/C	C	C/T	G/C	G	C/T	T/G	C/T

Table 4 (Part 3). Genotypes and Haplotype Pairs Observed for POR Gene

Genotype Number	HAP Pair		Polymorphic Sites					
			PS21	PS22	PS23	PS24	PS25	PS26
1	8	8	C	C	G	C	G	C
2	10	10	T	C	G	C	G	C
3	19	19	C	C	A	C	G	C
4	15	15	C	C	G	C	G	C
5	8	11	C/T	C	G	C	G	C
6	8	3	C/T	C	G	C	G	C/T
7	19	21	C	C	A	C	G	C
8	10	17	T/C	C	G	C	G	C
9	15	2	C/T	C	G	C	G	C
10	19	14	C	C	A/G	C	G	C
11	8	16	C	C	G	C	G	C
12	8	20	C	C	G/A	C	G	C
13	19	2	C/T	C	A/G	C	G	C
14	19	4	C	C	A	C	G	C
15	14	9	C/T	C	G/A	C	G	C
16	8	13	C/T	C	G	C	G	C
17	8	12	C/T	C	G	C	G	C
18	8	1	C	C	G	C	G	C
19	8	6	C/T	C/T	G	C	G	C
20	8	15	C	C	G	C	G	C
21	8	17	C	C	G	C	G	C
22	8	7	C	C	G	C	G/A	C
23	10	5	T/C	C	G/A	C	G	C
24	8	2	C/T	C	G	C	G	C
25	8	19	C	C	G/A	C	G	C
26	10	15	T/C	C	G	C	G	C
27	8	10	C/T	C	G	C	G	C
28	10	19	T/C	C	G/A	C	G	C
29	19	17	C	C	A/G	C	G	C
30	15	18	C	C	G	C	G	C
31	19	15	C	C	A/G	C	G	C

The haplotype pairs shown in Table 4 were estimated from the unphased genotypes using a computer-implemented extension of Clark's algorithm (Clark, A.G. 1990 *Mol Bio Evol* 7, 111-122) for assigning haplotypes to unrelated individuals in a population sample, as described in .PCT/US01/12831, filed April 18, 2001. In this method, haplotypes are assigned directly from individuals who are homozygous at all sites or heterozygous at no more than one of the variable sites. This list of haplotypes is then used to deconvolute the unphased genotypes in the remaining (multiply heterozygous) individuals. In the present analysis, the list of haplotypes was augmented with haplotypes obtained from two families (one three-generation Caucasian family and one two-generation African-American family).

By following this protocol, it was determined that the Index Repository examined herein and, by extension, the general population contains the 21 human POR haplotypes shown in Table 5 below.

A POR isogene defined by a full-haplotype shown in Table 5 below comprises the regions of the SEQ ID NOS indicated in Table 5, with their corresponding set of polymorphic locations and

identities, which are also set forth in Table 5.

Table 5 (Part 1). Haplotypes of the POR gene.

Regions Examined(a)	PS No.(b)	PS Position(c)	Haplotype Number(d)									
			1	2	3	4	5	6	7	8	9	10
3878-4429	-	-	-	-	-	-	-	-	-	-	-	-
22306-22748	1	22635/30	A	A	A	A	G	G	G	G	G	G
29363-29820	2	29577/150	C	C	C	C	C	C	C	C	C	C
29363-29820	3	29754/270	C	C	C	T	C	C	C	C	C	C
30215-30747	4	30427/390	C	C	C	C	C	C	C	C	C	C
30942-31856	5	31531/510	A	A	A	A	A	A	A	A	A	A
30942-31856	6	31644/630	C	C	C	C	C	C	C	C	C	C
30942-31856	7	31648/750	C	C	C	C	C	C	C	C	C	C
32039-32600	8	32446/870	A	A	A	A	A	A	A	A	A	A
32039-32600	9	32478/990	C	C	C	C	C	C	C	C	C	C
33448-34167	10	33538/1110	G	G	G	G	G	G	G	G	G	G
33448-34167	11	33551/1230	A	A	A	A	A	A	A	A	A	A
33448-34167	12	33571/1350	C	T	T	C	C	C	C	C	T	T
34661-36857	13	34766/1470	T	C	C	T	C	C	T	T	C	C
34661-36857	14	34796/1590	C	C	C	C	C	C	C	C	C	C
34661-36857	15	34797/1710	C	C	C	C	C	C	C	C	C	C
34661-36857	16	34850/1830	G	G	G	G	G	G	G	G	G	G
34661-36857	17	35325/1950	G	A	A	G	G	G	G	G	G	G
34661-36857	18	35631/2070	C	C	T	C	C	C	C	C	C	C
34661-36857	19	35632/2190	T	T	T	T	T	T	T	T	T	T
34661-36857	20	35721/2310	C	C	C	C	C	C	C	C	C	C
34661-36857	21	35774/2430	C	T	T	C	C	T	C	C	T	T
34661-36857	22	35992/2550	C	C	C	C	C	T	C	C	C	C
34661-36857	23	36054/2670	G	G	G	A	A	G	G	G	A	G
34661-36857	24	36207/2790	C	C	C	C	C	C	C	C	C	C
34661-36857	25	36330/2910	G	G	G	G	G	G	A	G	G	G
34661-36857	26	36617/3030	C	C	T	C	C	C	C	C	C	C

(a) Region examined represents the nucleotide positions defining the start and stop positions within SEQ ID NO:1 of the regions sequenced;

(b) PS = polymorphic site;

(c) Position of PS within the indicated SEQ ID NO, with the 1st position number referring to SEQ ID NO:1 and the 2nd position number referring to SEQ ID NO:134, a modified version of SEQ ID NO:1 that comprises the context sequence of each polymorphic site, PS1-PS26, to facilitate electronic searching of the haplotypes;

(d) Alleles for POR haplotypes are presented 5′ to 3′ in each column.

Table 5 (Part 2). Haplotypes of the POR gene.

Regions Examined(a)	PS No.(b)	PS Position(c)	Haplotype Number(d)									
			11	12	13	14	15	16	17	18	19	20
3878-4429	-	-	-	-	-	-	-	-	-	-	-	-
22306-22748	1	22635/30	G	G	G	G	G	G	G	G	G	G
29363-29820	2	29577/150	C	C	C	C	C	C	C	C	C	C
29363-29820	3	29754/270	C	C	C	C	C	C	C	C	T	T
30215-30747	4	30427/390	C	C	C	C	C	C	C	T	C	C
30942-31856	5	31531/510	A	A	A	G	G	G	G	G	A	A
30942-31856	6	31644/630	C	C	C	C	C	C	C	C	C	C
30942-31856	7	31648/750	C	C	T	C	C	C	C	C	C	C
32039-32600	8	32446/870	A	G	A	A	A	A	A	A	A	A
32039-32600	9	32478/990	C	C	C	C	C	C	C	C	C	T
33448-34167	10	33538/1110	G	G	G	C	C	C	C	C	G	G
33448-34167	11	33551/1230	A	A	A	G	G	G	G	G	A	A
33448-34167	12	33571/1350	T	T	T	C	C	C	C	C	C	C
34661-36857	13	34766/1470	T	C	C	C	C	C	C	C	T	T
34661-36857	14	34796/1590	C	C	C	C	C	C	T	C	C	C
34661-36857	15	34797/1710	C	C	C	T	T	T	T	T	C	C
34661-36857	16	34850/1830	G	G	G	C	C	G	C	C	G	G
34661-36857	17	35325/1950	G	G	G	G	G	G	G	G	G	G
34661-36857	18	35631/2070	C	C	C	T	T	T	T	T	C	C
34661-36857	19	35632/2190	T	T	T	G	G	G	G	G	T	T
34661-36857	20	35721/2310	C	C	C	C	T	T	T	C	C	C
34661-36857	21	35774/2430	T	T	T	C	C	C	C	C	C	C
34661-36857	22	35992/2550	C	C	C	C	C	C	C	C	C	C
34661-36857	23	36054/2670	G	G	G	G	G	G	G	G	A	A
34661-36857	24	36207/2790	C	C	C	C	C	C	C	C	C	C
34661-36857	25	36330/2910	G	G	G	G	G	G	G	G	G	G
34661-36857	26	36617/3030	C	C	C	C	C	C	C	C	C	C

(a) Region examined represents the nucleotide positions defining the start and stop positions within SEQ ID NO:1 of the regions sequenced;

(b) PS = polymorphic site;

(c) Position of PS within the indicated SEQ ID NO, with the 1st position number referring to SEQ ID NO:1 and the 2nd position number referring to SEQ ID NO:134, a modified version of SEQ ID NO:1 that comprises the context sequence of each polymorphic site, PS1-PS26, to facilitate electronic searching of the haplotypes;

(d) Alleles for POR haplotypes are presented 5′ to 3′ in each column.

Table 5 (Part 3). Haplotypes of the POR gene.

	Regions Examined(a)	PS No.(b)	PS Position(c)	Haplotype Number(d) 21
5	3878-4429	-	-	-
	22306-22748	1	22635/30	G
	29363-29820	2	29577/150	T
	29363-29820	3	29754/270	C
	30215-30747	4	30427/390	C
10	30942-31856	5	31531/510	A
	30942-31856	6	31644/630	C
	30942-31856	7	31648/750	C
	32039-32600	8	32446/870	A
	32039-32600	9	32478/990	C
15	33448-34167	10	33538/1110	G
	33448-34167	11	33551/1230	A
	33448-34167	12	33571/1350	C
	34661-36857	13	34766/1470	T
	34661-36857	14	34796/1590	C
20	34661-36857	15	34797/1710	C
	34661-36857	16	34850/1830	G
	34661-36857	17	35325/1950	G
	34661-36857	18	35631/2070	C
	34661-36857	19	35632/2190	T
25	34661-36857	20	35721/2310	C
	34661-36857	21	35774/2430	C
	34661-36857	22	35992/2550	C
	34661-36857	23	36054/2670	A
	34661-36857	24	36207/2790	C
30	34661-36857	25	36330/2910	G
	34661-36857	26	36617/3030	C

(a) Region examined represents the nucleotide positions defining the start and stop positions within SEQ ID NO:1 of the regions sequenced;

35   (b) PS = polymorphic site;

(c) Position of PS within the indicated SEQ ID NO, with the 1st position number referring to SEQ ID NO:1 and the 2nd position number referring to SEQ ID NO:134, a modified version of SEQ ID NO:1 that comprises the context sequence of each polymorphic site, PS1-PS26, to facilitate electronic searching of the haplotypes;

40   (d) Alleles for POR haplotypes are presented 5' to 3' in each column.

SEQ ID NO:1 refers to Figure 1, with the two alternative allelic variants of each polymorphic site indicated by the appropriate nucleotide symbol. SEQ ID NO:134 is a modified version of SEQ ID NO:1 that shows the context sequence of each of PS1-PS26 in a uniform format to facilitate electronic

45   searching of the POR haplotypes. For each polymorphic site, SEQ ID NO:134 contains a block of 60 bases of the nucleotide sequence encompassing the centrally-located polymorphic site at the 30th position, followed by 60 bases of unspecified sequence to represent that each polymorphic site is separated by genomic sequence whose composition is defined elsewhere herein.

Table 6 below shows the percent of chromosomes characterized by a given POR haplotype for

50   all unrelated individuals in the Index Repository for which haplotype data was obtained. The percent

of these unrelated individuals who have a given POR haplotype pair is shown in Table 7. In Tables 6 and 7, the "Total" column shows this frequency data for all of these unrelated individuals, while the other columns show the frequency data for these unrelated individuals categorized according to their self-identified ethnogeographic origin. Abbreviations used in Tables 6 and 7 are AF = African

5   Descent, AS = Asian, CA = Caucasian, HL = Hispanic-Latino, and AM = Native American.

Table 6. Frequency of Observed POR Haplotypes In Unrelated Individuals

	HAP No.	HAP ID	Total	CA	AF	AS	HL	AM
10	1	6670644	0.62	2.5	0.0	0.0	0.0	0.0
	2	6670633	4.94	2.5	5.0	0.0	11.11	16.67
	3	6670639	0.62	0.0	0.0	0.0	2.78	0.0
	4	6670642	0.62	0.0	0.0	0.0	2.78	0.0
	5	6670643	0.62	0.0	2.5	0.0	0.0	0.0
15	6	6670640	0.62	0.0	2.5	0.0	0.0	0.0
	7	6670673	0.62	0.0	2.5	0.0	0.0	0.0
	8	6670629	32.72	37.5	27.5	32.5	38.89	0.0
	9	6670636	0.62	2.5	0.0	0.0	0.0	0.0
	10	6670630	17.9	7.5	7.5	42.5	16.67	0.0
20	11	6670637	0.62	0.0	0.0	0.0	2.78	0.0
	12	6670653	0.62	2.5	0.0	0.0	0.0	0.0
	13	6670646	0.62	2.5	0.0	0.0	0.0	0.0
	14	6670635	1.85	2.5	2.5	0.0	2.78	0.0
	15	6670632	12.35	7.5	17.5	17.5	5.56	16.67
25	16	6670684	0.62	0.0	0.0	2.5	0.0	0.0
	17	6670634	3.7	0.0	12.5	0.0	2.78	0.0
	18	6670638	0.62	0.0	2.5	0.0	0.0	0.0
	19	6670631	17.9	30.0	17.5	5.0	13.89	50.0
	20	6670645	0.62	2.5	0.0	0.0	0.0	0.0
30	21	6670641	0.62	0.0	0.0	0.0	0.0	16.67

Table 7. Frequency of Observed POR Haplotype Pairs In Unrelated Individuals

HAP1	HAP2	Total	CA	AF	AS	HL	AM
8	8	11.11	10.0	10.0	15.0	11.11	0.0
10	10	4.94	0.0	0.0	15.0	5.56	0.0
19	19	2.47	10.0	0.0	0.0	0.0	0.0
15	15	2.47	0.0	5.0	5.0	0.0	0.0
8	11	1.23	0.0	0.0	0.0	5.56	0.0
8	3	1.23	0.0	0.0	0.0	5.56	0.0
19	21	1.23	0.0	0.0	0.0	0.0	33.33
10	17	1.23	0.0	5.0	0.0	0.0	0.0
15	2	1.23	0.0	5.0	0.0	0.0	0.0
19	14	2.47	0.0	5.0	0.0	5.56	0.0
8	16	1.23	0.0	0.0	5.0	0.0	0.0
8	20	1.23	5.0	0.0	0.0	0.0	0.0
19	2	3.7	5.0	0.0	0.0	5.56	33.33
19	4	1.23	0.0	0.0	0.0	5.56	0.0
14	9	1.23	5.0	0.0	0.0	0.0	0.0
8	13	1.23	5.0	0.0	0.0	0.0	0.0
8	12	1.23	5.0	0.0	0.0	0.0	0.0
8	1	1.23	5.0	0.0	0.0	0.0	0.0
8	6	1.23	0.0	5.0	0.0	0.0	0.0
8	15	4.94	5.0	5.0	5.0	5.56	0.0
8	17	3.7	0.0	10.0	0.0	5.56	0.0
8	7	1.23	0.0	5.0	0.0	0.0	0.0
10	5	1.23	0.0	5.0	0.0	0.0	0.0
8	2	4.94	0.0	5.0	0.0	16.67	0.0
8	19	7.41	20.0	5.0	0.0	5.56	0.0
10	15	6.17	0.0	0.0	20.0	5.56	0.0
8	10	11.11	10.0	0.0	25.0	11.11	0.0
10	19	6.17	5.0	5.0	10.0	5.56	0.0
19	17	2.47	0.0	10.0	0.0	0.0	0.0
15	18	1.23	0.0	5.0	0.0	0.0	0.0
19	15	6.17	10.0	10.0	0.0	0.0	33.33

The size and composition of the Index Repository were chosen to represent the genetic diversity across and within four major population groups comprising the general United States population. For example, as described in Table 1 above, this repository contains approximately equal sample sizes of African-descent, Asian-American, European-American, and Hispanic-Latino population groups. Almost all individuals representing each group had all four grandparents with the same ethnogeographic background. The number of unrelated individuals in the Index Repository provides a sample size that is sufficient to detect SNPs and haplotypes that occur in the general population with high statistical certainty. For instance, a haplotype that occurs with a frequency of 5% in the general population has a probability higher than 99.9% of being observed in a sample of 80 individuals from the general population. Similarly, a haplotype that occurs with a frequency of 10% in a specific population group has a 99% probability of being observed in a sample of 20 individuals from that population group. In addition, the size and composition of the Index Repository means that the relative frequencies determined therein for the haplotypes and haplotype pairs of the POR gene are likely to be similar to the relative frequencies of these POR haplotypes and haplotype pairs in the

general U.S. population and in the four population groups represented in the Index Repository. The

genetic diversity observed for the three Native Americans is presented because it is of scientific

interest, but due to the small sample size it lacks statistical significance.

In view of the above, it will be seen that the several advantages of the invention are achieved

5   and other advantageous results attained.

As various changes could be made in the above methods and compositions without departing

from the scope of the invention, it is intended that all matter contained in the above description and

shown in the accompanying drawings shall be interpreted as illustrative and not in a limiting sense.

All references cited in this specification, including patents and patent applications, are hereby

10   incorporated in their entirety by reference. The discussion of references herein is intended merely to

summarize the assertions made by their authors and no admission is made that any reference

constitutes prior art. Applicants reserve the right to challenge the accuracy and pertinency of the cited

references.

What is Claimed is:

1. A method for haplotyping the P450 (cytochrome) oxidoreductase (POR) gene of an individual, which comprises determining which of the POR haplotypes shown in the table immediately below defines one copy of the individual's POR gene, wherein the determining step comprises identifying the phased sequence of nucleotides present at each of PS1-PS26 on at least one copy of the individual's POR gene, and wherein each of the POR haplotypes comprises a sequence of polymorphisms whose positions and identities are set forth in the table immediately below:

PS No.(a)	PS Position(b)	Haplotype Number(c) (Part 1)									
		1	2	3	4	5	6	7	8	9	10
1	22635	A	A	A	A	G	G	G	G	G	G
2	29577	C	C	C	C	C	C	C	C	C	C
3	29754	C	C	C	T	C	C	C	C	C	C
4	30427	C	C	C	C	C	C	C	C	C	C
5	31531	A	A	A	A	A	A	A	A	A	A
6	31644	C	C	C	C	C	C	C	C	C	C
7	31648	C	C	C	C	C	C	C	C	C	C
8	32446	A	A	A	A	A	A	A	A	A	A
9	32478	C	C	C	C	C	C	C	C	C	C
10	33538	G	G	G	G	G	G	G	G	G	G
11	33551	A	A	A	A	A	A	A	A	A	A
12	33571	C	T	T	C	C	C	C	C	T	T
13	34766	T	C	C	T	C	C	T	T	C	C
14	34796	C	C	C	C	C	C	C	C	C	C
15	34797	C	C	C	C	C	C	C	C	C	C
16	34850	G	G	G	G	G	G	G	G	G	G
17	35325	G	A	A	G	G	G	G	G	G	G
18	35631	C	C	T	C	C	C	C	C	C	C
19	35632	T	T	T	T	T	T	T	T	T	T
20	35721	C	C	C	C	C	C	C	C	C	C
21	35774	C	T	T	C	C	T	C	C	T	T
22	35992	C	C	C	C	C	T	C	C	C	C
23	36054	G	G	G	A	A	G	G	G	A	G
24	36207	C	C	C	C	C	C	C	C	C	C
25	36330	G	G	G	G	G	G	A	G	G	G
26	36617	C	C	T	C	C	C	C	C	C	C

(a) PS = polymorphic site;
(b) Position of PS within SEQ ID NO:1;
(c) Alleles for haplotypes are presented 5′ to 3′ in each column;

PS No.(a)	PS Position(b)	Haplotype Number(c) (Part 2)									
		11	12	13	14	15	16	17	18	19	20
1	22635	G	G	G	G	G	G	G	G	G	G
2	29577	C	C	C	C	C	C	C	C	C	C
3	29754	C	C	C	C	C	C	C	C	T	T
4	30427	C	C	C	C	C	C	C	T	C	C
5	31531	A	A	A	G	G	G	G	G	A	A
6	31644	C	C	C	C	C	C	C	C	C	C
7	31648	C	C	T	C	C	C	C	C	C	C
8	32446	A	G	A	A	A	A	A	A	A	A
9	32478	C	C	C	C	C	C	C	C	C	T
10	33538	G	G	G	C	C	C	C	C	G	G
11	33551	A	A	A	G	G	G	G	G	A	A
12	33571	T	T	T	C	C	C	C	C	C	C
13	34766	T	C	C	C	C	C	C	C	T	T
14	34796	C	C	C	C	C	C	T	C	C	C
15	34797	C	C	C	T	T	T	T	T	C	C
16	34850	G	G	G	C	C	G	C	C	G	G
17	35325	G	G	G	G	G	G	G	G	G	G
18	35631	C	C	C	T	T	T	T	T	C	C
19	35632	T	T	T	G	G	G	G	G	T	T
20	35721	C	C	C	C	T	T	T	C	C	C
21	35774	T	T	T	C	C	C	C	C	C	C
22	35992	C	C	C	C	C	C	C	C	C	C
23	36054	G	G	G	G	G	G	G	G	A	A
24	36207	C	C	C	C	C	C	C	C	C	C
25	36330	G	G	G	G	G	G	G	G	G	G
26	36617	C	C	C	C	C	C	C	C	C	C

(a) PS = polymorphic site;
(b) Position of PS within SEQ ID NO:1;
(c) Alleles for haplotypes are presented 5´ to 3´ in each column;

PS No.(a)	PS Position(b)	Haplotype Number(c) (Part 3) 21
1	22635	G
2	29577	T
3	29754	C
4	30427	C
5	31531	A
6	31644	C
7	31648	C
8	32446	A
9	32478	C
10	33538	G
11	33551	A
12	33571	C
13	34766	T
14	34796	C
15	34797	C
16	34850	G
17	35325	G
18	35631	C
19	35632	T
20	35721	C
21	35774	C
22	35992	C
23	36054	A
24	36207	C
25	36330	G
26	36617	C

(a) PS = polymorphic site;
(b) Position of PS within SEQ ID NO:1;
(c) Alleles for haplotypes are presented 5′ to 3′ in each column.

2.   A method for haplotyping the P450 (cytochrome) oxidoreductase (POR) gene of an individual,
which comprises determining which of the POR haplotype pairs shown in the table
immediately below defines both copies of the individual's POR gene, wherein the determining
step comprises identifying the phased sequence of nucleotides present at each of PS1-PS26 on
both copies of the individual's POR gene, and wherein each of the POR haplotype pairs
consists of first and second haplotypes which comprise first and second sequences of
polymorphisms whose positions and identities are set forth in the table immediately below:

	PS No.(a)	PS Position(b)	Haplotype Pair(c) (Part 1)							
			8/8	10/10	19/19	15/15	8/11	8/3	19/21	10/17
5	1	22635	G/G	G/G	G/G	G/G	G/G	G/A	G/G	G/G
	2	29577	C/C	C/C	C/C	C/C	C/C	C/C	C/T	C/C
	3	29754	C/C	C/C	T/T	C/C	C/C	C/C	T/C	C/C
	4	30427	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
10	5	31531	A/A	A/A	A/A	G/G	A/A	A/A	A/A	A/G
	6	31644	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	7	31648	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	8	32446	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
	9	32478	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
15	10	33538	G/G	G/G	G/G	C/C	G/G	G/G	G/G	G/C
	11	33551	A/A	A/A	A/A	G/G	A/A	A/A	A/A	A/G
	12	33571	C/C	T/T	C/C	C/C	C/T	C/T	C/C	T/C
	13	34766	T/T	C/C	T/T	C/C	T/T	T/C	T/T	C/C
	14	34796	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/T
20	15	34797	C/C	C/C	C/C	T/T	C/C	C/C	C/C	C/T
	16	34850	G/G	G/G	G/G	C/C	G/G	G/G	G/G	G/C
	17	35325	G/G	G/G	G/G	G/G	G/G	G/A	G/G	G/G
	18	35631	C/C	C/C	C/C	T/T	C/C	C/T	C/C	C/T
	19	35632	T/T	T/T	T/T	G/G	T/T	T/T	T/T	T/G
25	20	35721	C/C	C/C	C/C	T/T	C/C	C/C	C/C	C/T
	21	35774	C/C	T/T	C/C	C/C	C/T	C/T	C/C	T/C
	22	35992	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	23	36054	G/G	G/G	A/A	G/G	G/G	G/G	A/A	G/G
	24	36207	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
30	25	36330	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
	26	36617	C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/C

(a) PS = polymorphic site;

(b) Position of PS in SEQ ID NO:1;

(c) Haplotype pairs are represented as 1st haplotype/2nd haplotype; with alleles of each haplotype shown 5′ to 3′ as 1st polymorphism/2nd polymorphism in each column;

PS No.(a)	PS Position(b)	Haplotype Pair(c) (Part 2)							
		15/2	19/14	8/16	8/20	19/2	19/4	14/9	8/13
1	22635	G/A	G/G	G/G	G/G	G/A	G/A	G/G	G/G
2	29577	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
3	29754	C/C	T/C	C/C	C/T	T/C	T/T	C/C	C/C
4	30427	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
5	31531	G/A	A/G	A/G	A/A	A/A	A/A	G/A	A/A
6	31644	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
7	31648	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/T
8	32446	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
9	32478	C/C	C/C	C/C	C/T	C/C	C/C	C/C	C/C
10	33538	C/G	G/C	G/C	G/G	G/G	G/G	C/G	G/G
11	33551	G/A	A/G	A/G	A/A	A/A	A/A	G/A	A/A
12	33571	C/T	C/C	C/C	C/C	C/T	C/C	C/T	C/T
13	34766	C/C	T/C	T/C	T/T	T/C	T/T	C/C	T/C
14	34796	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
15	34797	T/C	C/T	C/T	C/C	C/C	C/C	T/C	C/C
16	34850	C/G	G/C	G/G	G/G	G/G	G/G	C/G	G/G
17	35325	G/A	G/G	G/G	G/G	G/A	G/G	G/G	G/G
18	35631	T/C	C/T	C/T	C/C	C/C	C/C	T/C	C/C
19	35632	G/T	T/G	T/G	T/T	T/T	T/T	G/T	T/T
20	35721	T/C	C/C	C/T	C/C	C/C	C/C	C/C	C/C
21	35774	C/T	C/C	C/C	C/C	C/T	C/C	C/T	C/T
22	35992	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
23	36054	G/G	A/G	G/G	G/A	A/G	A/A	G/A	G/G
24	36207	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
25	36330	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
26	36617	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C

(a) PS = polymorphic site;

(b) Position of PS in SEQ ID NO:1;

(c) Haplotype pairs are represented as 1st haplotype/2nd haplotype; with alleles of each haplotype shown 5' to 3' as 1st polymorphism/2nd polymorphism in each column;

PS No.(a)	PS Position(b)	Haplotype Pair(c) (Part 3)							
		8/12	8/1	8/6	8/15	8/17	8/7	10/5	8/2
1	22635	G/G	G/A	G/G	G/G	G/G	G/G	G/G	G/A
2	29577	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
3	29754	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
4	30427	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
5	31531	A/A	A/A	A/A	A/G	A/G	A/A	A/A	A/A
6	31644	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
7	31648	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
8	32446	A/G	A/A	A/A	A/A	A/A	A/A	A/A	A/A
9	32478	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
10	33538	G/G	G/G	G/G	G/C	G/C	G/G	G/G	G/G
11	33551	A/A	A/A	A/A	A/G	A/G	A/A	A/A	A/A
12	33571	C/T	C/C	C/C	C/C	C/C	C/C	T/C	C/T
13	34766	T/C	T/T	T/C	T/C	T/C	T/T	C/C	T/C
14	34796	C/C	C/C	C/C	C/C	C/T	C/C	C/C	C/C
15	34797	C/C	C/C	C/C	C/T	C/T	C/C	C/C	C/C
16	34850	G/G	G/G	G/G	G/C	G/C	G/G	G/G	G/G
17	35325	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/A
18	35631	C/C	C/C	C/C	C/T	C/T	C/C	C/C	C/C
19	35632	T/T	T/T	T/T	T/G	T/G	T/T	T/T	T/T
20	35721	C/C	C/C	C/C	C/T	C/T	C/C	C/C	C/C
21	35774	C/T	C/C	C/T	C/C	C/C	C/C	T/C	C/T
22	35992	C/C	C/C	C/T	C/C	C/C	C/C	C/C	C/C
23	36054	G/G	G/G	G/G	G/G	G/G	G/G	G/A	G/G
24	36207	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
25	36330	G/G	G/G	G/G	G/G	G/G	G/A	G/G	G/G
26	36617	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C

(a) PS = polymorphic site;

(b) Position of PS in SEQ ID NO:1;

(c) Haplotype pairs are represented as 1st haplotype/2nd haplotype; with alleles of each haplotype shown 5′ to 3′ as 1st polymorphism/2nd polymorphism in each column;

PS No.(a)	PS Position(b)	Haplotype Pair(c) (Part 4)						
		8/19	10/15	8/10	10/19	19/17	15/18	19/15
1	22635	G/G	G/G	G/G	G/G	G/G	G/G	G/G
2	29577	C/C	C/C	C/C	C/C	C/C	C/C	C/C
3	29754	C/T	C/C	C/C	C/T	T/C	C/C	T/C
4	30427	C/C	C/C	C/C	C/C	C/C	C/T	C/C
5	31531	A/A	A/G	A/A	A/A	A/G	G/G	A/G
6	31644	C/C	C/C	C/C	C/C	C/C	C/C	C/C
7	31648	C/C	C/C	C/C	C/C	C/C	C/C	C/C
8	32446	A/A	A/A	A/A	A/A	A/A	A/A	A/A
9	32478	C/C	C/C	C/C	C/C	C/C	C/C	C/C
10	33538	G/G	G/C	G/G	G/G	G/C	C/C	G/C
11	33551	A/A	A/G	A/A	A/A	A/G	G/G	A/G
12	33571	C/C	T/C	C/T	T/C	C/C	C/C	C/C
13	34766	T/T	C/C	T/C	C/T	T/C	C/C	T/C
14	34796	C/C	C/C	C/C	C/C	C/T	C/C	C/C
15	34797	C/C	C/T	C/C	C/C	C/T	T/T	C/T
16	34850	G/G	G/C	G/G	G/G	G/C	C/C	G/C
17	35325	G/G	G/G	G/G	G/G	G/G	G/G	G/G
18	35631	C/C	C/T	C/C	C/C	C/T	T/T	C/T
19	35632	T/T	T/G	T/T	T/T	T/G	G/G	T/G
20	35721	C/C	C/T	C/C	C/C	C/T	T/C	C/T
21	35774	C/C	T/C	C/T	T/C	C/C	C/C	C/C
22	35992	C/C	C/C	C/C	C/C	C/C	C/C	C/C
23	36054	G/A	G/G	G/G	G/A	A/G	G/G	A/G
24	36207	C/C	C/C	C/C	C/C	C/C	C/C	C/C
25	36330	G/G	G/G	G/G	G/G	G/G	G/G	G/G
26	36617	C/C	C/C	C/C	C/C	C/C	C/C	C/C

(a) PS = polymorphic site;

(b) Position of PS in SEQ ID NO:1;

(c) Haplotype pairs are represented as 1st haplotype/2nd haplotype; with alleles of each haplotype shown 5' to 3' as 1st polymorphism/2nd polymorphism in each column.

3. A method for genotyping the P450 (cytochrome) oxidoreductase (POR) gene of an individual, comprising determining for the two copies of the POR gene present in the individual the identity of the nucleotide pair at one or more polymorphic sites (PS) selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18, PS19, PS20, PS21, PS22, PS23, PS24, PS25 and PS26, wherein the one or more polymorphic sites (PS) have the position and alternative alleles shown in SEQ ID NO:1.

4. The method of claim 3, wherein the determining step comprises:

(a) isolating from the individual a nucleic acid mixture comprising both copies of the POR gene, or a fragment thereof, that are present in the individual;

(b) amplifying from the nucleic acid mixture a target region containing one of the selected polymorphic sites;

(c) hybridizing a primer extension oligonucleotide to one allele of the amplified target region, wherein the oligonucleotide is designed for genotyping the selected polymorphic

site in the target region;

(d)   performing a nucleic acid template-dependent, primer extension reaction on the

10   hybridized oligonucleotide in the presence of at least one terminator of the reaction,

wherein the terminator is complementary to one of the alternative nucleotides present at

the selected polymorphic site; and

(e)   detecting the presence and identity of the terminator in the extended oligonucleotide.

5.   The method of claim 3, which comprises determining for the two copies of the POR gene
present in the individual the identity of the nucleotide pair at each of PS1-PS26.

6.   A method for haplotyping the P450 (cytochrome) oxidoreductase (POR) gene of an individual
which comprises determining, for one copy of the POR gene present in the individual, the
identity of the nucleotide at two or more polymorphic sites (PS) selected from the group
consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14,
PS15, PS16, PS17, PS18, PS19, PS20, PS21, PS22, PS23, PS24, PS25 and PS26, wherein the
selected PS have the position and alternative alleles shown in SEQ ID NO:1.

7.   The method of claim 6, wherein the determining step comprises:

(a)   isolating from the individual a nucleic acid sample containing only one of the two copies

of the POR gene, or a fragment thereof, that is present in the individual;

(b)   amplifying from the nucleic acid sample a target region containing one of the selected

5   polymorphic sites;

(c)   hybridizing a primer extension oligonucleotide to one allele of the amplified target region,

wherein the oligonucleotide is designed for haplotyping the selected polymorphic site in

the target region;

(d)   performing a nucleic acid template-dependent, primer extension reaction on the

10   hybridized oligonucleotide in the presence of at least one terminator of the reaction,

wherein the terminator is complementary to one of the alternative nucleotides present at

the selected polymorphic site; and

(e)   detecting the presence and identity of the terminator in the extended oligonucleotide.

8.   A method for predicting a haplotype pair for the P450 (cytochrome) oxidoreductase (POR) gene
of an individual comprising:

(a)   identifying a POR genotype for the individual, wherein the genotype comprises the

nucleotide pair at two or more polymorphic sites (PS) selected from the group consisting

5   of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15,

PS16, PS17, PS18, PS19, PS20, PS21, PS22, PS23, PS24, PS25 and PS26, wherein the

selected PS have the position and alternative alleles shown in SEQ ID NO:1;

(b)   comparing the genotype to the haplotype pair data set forth in the table immediately

below; and

10   (c)   determining which haplotype pair is consistent with the genotype of the individual and

with the haplotype pair data

PS No.(a)	PS Position(b)	Haplotype Pair(c) (Part 1)							
		8/8	10/10	19/19	15/15	8/11	8/3	19/21	10/17
1	22635	G/G	G/G	G/G	G/G	G/G	G/A	G/G	G/G
2	29577	C/C	C/C	C/C	C/C	C/C	C/C	C/T	C/C
3	29754	C/C	C/C	T/T	C/C	C/C	C/C	T/C	C/C
4	30427	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
5	31531	A/A	A/A	A/A	G/G	A/A	A/A	A/A	A/G
6	31644	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
7	31648	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
8	32446	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
9	32478	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
10	33538	G/G	G/G	G/G	C/C	G/G	G/G	G/G	G/C
11	33551	A/A	A/A	A/A	G/G	A/A	A/A	A/A	A/G
12	33571	C/C	T/T	C/C	C/C	C/T	C/T	C/C	T/C
13	34766	T/T	C/C	T/T	C/C	T/T	T/C	T/T	C/C
14	34796	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/T
15	34797	C/C	C/C	C/C	T/T	C/C	C/C	C/C	C/T
16	34850	G/G	G/G	G/G	C/C	G/G	G/G	G/G	G/C
17	35325	G/G	G/G	G/G	G/G	G/G	G/A	G/G	G/G
18	35631	C/C	C/C	C/C	T/T	C/C	C/T	C/C	C/T
19	35632	T/T	T/T	T/T	G/G	T/T	T/T	T/T	T/G
20	35721	C/C	C/C	C/C	T/T	C/C	C/C	C/C	C/T
21	35774	C/C	T/T	C/C	C/C	C/T	C/T	C/C	T/C
22	35992	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
23	36054	G/G	G/G	A/A	G/G	G/G	G/G	A/A	G/G
24	36207	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
25	36330	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
26	36617	C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/C

(a) PS = polymorphic site;
(b) Position of PS in SEQ ID NO:1;
(c) Haplotype pairs are represented as 1st haplotype/2nd haplotype; with alleles of each haplotype shown 5′ to 3′ as 1st polymorphism/2nd polymorphism in each column;

PS No.(a)	PS Position(b)	Haplotype Pair(c) (Part 2)							
		15/2	19/14	8/16	8/20	19/2	19/4	14/9	8/13
1	22635	G/A	G/G	G/G	G/G	G/A	G/A	G/G	G/G
2	29577	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
3	29754	C/C	T/C	C/C	C/T	T/C	T/T	C/C	C/C
4	30427	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
5	31531	G/A	A/G	A/G	A/A	A/A	A/A	G/A	A/A
6	31644	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
7	31648	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/T
8	32446	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
9	32478	C/C	C/C	C/C	C/T	C/C	C/C	C/C	C/C
10	33538	C/G	G/C	G/C	G/G	G/G	G/G	C/G	G/G
11	33551	G/A	A/G	A/G	A/A	A/A	A/A	G/A	A/A
12	33571	C/T	C/C	C/C	C/C	C/T	C/C	C/T	C/T
13	34766	C/C	T/C	T/C	T/T	T/C	T/T	C/C	T/C
14	34796	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
15	34797	T/C	C/T	C/T	C/C	C/C	C/C	T/C	C/C
16	34850	C/G	G/C	G/G	G/G	G/G	G/G	C/G	G/G
17	35325	G/A	G/G	G/G	G/G	G/A	G/G	G/G	G/G
18	35631	T/C	C/T	C/T	C/C	C/C	C/C	T/C	C/C
19	35632	G/T	T/G	T/G	T/T	T/T	T/T	G/T	T/T
20	35721	T/C	C/C	C/T	C/C	C/C	C/C	C/C	C/C
21	35774	C/T	C/C	C/C	C/C	C/T	C/C	C/T	C/T
22	35992	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
23	36054	G/G	A/G	G/G	G/A	A/G	A/A	G/A	G/G
24	36207	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
25	36330	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
26	36617	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C

(a) PS = polymorphic site;

(b) Position of PS in SEQ ID NO:1;

(c) Haplotype pairs are represented as 1st haplotype/2nd haplotype; with alleles of each haplotype shown 5' to 3' as 1st polymorphism/2nd polymorphism in each column;

PS No.(a)	PS Position(b)	Haplotype Pair(c) (Part 3)							
		8/12	8/1	8/6	8/15	8/17	8/7	10/5	8/2
1	22635	G/G	G/A	G/G	G/G	G/G	G/G	G/G	G/A
2	29577	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
3	29754	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
4	30427	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
5	31531	A/A	A/A	A/A	A/G	A/G	A/A	A/A	A/A
6	31644	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
7	31648	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
8	32446	A/G	A/A	A/A	A/A	A/A	A/A	A/A	A/A
9	32478	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
10	33538	G/G	G/G	G/G	G/C	G/C	G/G	G/G	G/G
11	33551	A/A	A/A	A/A	A/G	A/G	A/A	A/A	A/A
12	33571	C/T	C/C	C/C	C/C	C/C	C/C	T/C	C/T
13	34766	T/C	T/T	T/C	T/C	T/C	T/T	C/C	T/C
14	34796	C/C	C/C	C/C.	C/C	C/T	C/C	C/C	C/C
15	34797	C/C	C/C	C/C	C/T	C/T	C/C	C/C	C/C
16	34850	G/G	G/G	G/G	G/C	G/C	G/G	G/G	G/G
17	35325	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/A
18	35631	C/C	C/C	C/C	C/T	C/T	C/C	C/C	C/C
19	35632	T/T	T/T	T/T	T/G	T/G	T/T	T/T	T/T
20	35721	C/C	C/C	C/C	C/T	C/T	C/C	C/C	C/C
21	35774	C/T	C/C	C/T	C/C	C/C	C/C	T/C	C/T
22	35992	C/C	C/C	C/T	C/C	C/C	C/C	C/C	C/C
23	36054	G/G	G/G	G/G	G/G	G/G	G/G	G/A	G/G
24	36207	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
25	36330	G/G	G/G	G/G	G/G	G/G	G/A	G/G	G/G
26	36617	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C

(a) PS = polymorphic site;

(b) Position of PS in SEQ ID NO:1;

(c) Haplotype pairs are represented as 1st haplotype/2nd haplotype; with alleles of each haplotype shown 5′ to 3′ as 1st polymorphism/2nd polymorphism in each column;

PS No.(a)	PS Position(b)	Haplotype Pair(c) (Part 4)						
		8/19	10/15	8/10	10/19	19/17	15/18	19/15
1	22635	G/G	G/G	G/G	G/G	G/G	G/G	G/G
2	29577	C/C	C/C	C/C	C/C	C/C	C/C	C/C
3	29754	C/T	C/C	C/C	C/T	T/C	C/C	T/C
4	30427	C/C	C/C	C/C	C/C	C/C	C/T	C/C
5	31531	A/A	A/G	A/A	A/A	A/G	G/G	A/G
6	31644	C/C	C/C	C/C	C/C	C/C	C/C	C/C
7	31648	C/C	C/C	C/C	C/C	C/C	C/C	C/C
8	32446	A/A	A/A	A/A	A/A	A/A	A/A	A/A
9	32478	C/C	C/C	C/C	C/C	C/C	C/C	C/C
10	33538	G/G	G/C	G/G	G/G	G/C	C/C	G/C
11	33551	A/A	A/G	A/A	A/A	A/G	G/G	A/G
12	33571	C/C	T/C	C/T	T/C	C/C	C/C	C/C
13	34766	T/T	C/C	T/C	C/T	T/C	C/C	T/C
14	34796	C/C	C/C	C/C	C/C	C/T	C/C	C/C
15	34797	C/C	C/T	C/C	C/C	C/T	T/T	C/T
16	34850	G/G	G/C	G/G	G/G	G/C	C/C	G/C
17	35325	G/G	G/G	G/G	G/G	G/G	G/G	G/G
18	35631	C/C	C/T	C/C	C/C	C/T	T/T	C/T
19	35632	T/T	T/G	T/T	T/T	T/G	G/G	T/G
20	35721	C/C	C/T	C/C	C/C	C/T	T/C	C/T
21	35774	C/C	T/C	C/T	T/C	C/C	C/C	C/C
22	35992	C/C	C/C	C/C	C/C	C/C	C/C	C/C
23	36054	G/A	G/G	G/G	G/A	A/G	G/G	A/G
24	36207	C/C	C/C	C/C	C/C	C/C	C/C	C/C
25	36330	G/G	G/G	G/G	G/G	G/G	G/G	G/G
26	36617	C/C	C/C	C/C	C/C	C/C	C/C	C/C

(a) PS = polymorphic site;

(b) Position of PS in SEQ ID NO:1;

(c) Haplotype pairs are represented as 1st haplotype/2nd haplotype; with alleles of each haplotype shown 5' to 3' as 1st polymorphism/2nd polymorphism in each column.

9. The method of claim 8, wherein the identified genotype of the individual comprises the nucleotide pair at each of PS1-PS26, which have the position and alternative alleles shown in SEQ ID NO:1.

10. A method for identifying an association between a trait and at least one haplotype or haplotype pair of the P450 (cytochrome) oxidoreductase (POR) gene which comprises comparing the frequency of the haplotype or haplotype pair in a population exhibiting the trait with the frequency of the haplotype or haplotype pair in a reference population, wherein the haplotype is selected from haplotypes 1-21 shown in the table presented immediately below, wherein each of the haplotypes comprises a sequence of polymorphisms whose positions and identities are set forth in the table immediately below:

PS No.(a)	PS Position(b)	Haplotype Number(c) (Part 1)									
		1	2	3	4	5	6	7	8	9	10
1	22635	A	A	A	A	G	G	G	G	G	G
2	29577	C	C	C	C	C	C	C	C	C	C
3	29754	C	C	C	T	C	C	C	C	C	C
4	30427	C	C	C	C	C	C	C	C	C	C
5	31531	A	A	A	A	A	A	A	A	A	A
6	31644	C	C	C	C	C	C	C	C	C	C
7	31648	C	C	C	C	C	C	C	C	C	C
8	32446	A	A	A	A	A	A	A	A	A	A
9	32478	C	C	C	C	C	C	C	C	C	C
10	33538	G	G	G	G	G	G	G	G	G	G
11	33551	A	A	A	A	A	A	A	A	A	A
12	33571	C	T	T	C	C	C	C	C	T	T
13	34766	T	C	C	T	C	C	T	T	C	C
14	34796	C	C	C	C	C	C	C	C	C	C
15	34797	C	C	C	C	C	C	C	C	C	C
16	34850	G	G	G	G	G	G	G	G	G	G
17	35325	G	A	A	G	G	G	G	G	G	G
18	35631	C	C	T	C	C	C	C	C	C	C
19	35632	T	T	T	T	T	T	T	T	T	T
20	35721	C	C	C	C	C	C	C	C	C	C
21	35774	C	T	T	C	C	T	C	C	T	T
22	35992	C	C	C	C	C	T	C	C	C	C
23	36054	G	G	G	A	A	G	G	G	A	G
24	36207	C	C	C	C	C	C	C	C	C	C
25	36330	G	G	G	G	G	G	A	G	G	G
26	36617	C	C	T	C	C	C	C	C	C	C

(a) PS = polymorphic site;
(b) Position of PS within SEQ ID NO:1;
(c) Alleles for haplotypes are presented 5′ to 3′ in each column;

PS No.(a)	PS Position(b)	Haplotype Number(c) (Part 2)									
		11	12	13	14	15	16	17	18	19	20
1	22635	G	G	G	G	G	G	G	G	G	G
2	29577	C	C	C	C	C	C	C	C	C	C
3	29754	C	C	C	C	C	C	C	C	T	T
4	30427	C	C	C	C	C	C	C	T	C	C
5	31531	A	A	A	G	G	G	G	G	A	A
6	31644	C	C	C	C	C	C	C	C	C	C
7	31648	C	C	T	C	C	C	C	C	C	C
8	32446	A	G	A	A	A	A	A	A	A	A
9	32478	C	C	C	C	C	C	C	C	C	T
10	33538	G	G	G	C	C	C	C	C	G	G
11	33551	A	A	A	G	G	G	G	G	A	A
12	33571	T	T	T	C	C	C	C	C	C	C
13	34766	T	C	C	C	C	C	C	C	T	T
14	34796	C	C	C	C	C	C	T	C	C	C
15	34797	C	C	C	T	T	T	T	T	C	C
16	34850	G	G	G	C	C	G	C	C	G	G
17	35325	G	G	G	G	G	G	G	G	G	G
18	35631	C	C	C	T	T	T	T	T	C	C
19	35632	T	T	T	G	G	G	G	G	T	T
20	35721	C	C	C	C	T	T	T	C	C	C
21	35774	T	T	T	C	C	C	C	C	C	C
22	35992	C	C	C	C	C	C	C	C	C	C
23	36054	G	G	G	G	G	G	G	G	A	A
24	36207	C	C	C	C	C	C	C	C	C	C
25	36330	G	G	G	G	G	G	G	G	G	G
26	36617	C	C	C	C	C	C	C	C	C	C

(a) PS = polymorphic site;
(b) Position of PS within SEQ ID NO:1;
(c) Alleles for haplotypes are presented 5' to 3' in each column;

PS No.(a)	PS Position(b)	Haplotype Number(c) (Part 3) 21
1	22635	G
2	29577	T
3	29754	C
4	30427	C
5	31531	A
6	31644	C
7	31648	C
8	32446	A
9	32478	C
10	33538	G
11	33551	A
12	33571	C
13	34766	T
14	34796	C
15	34797	C
16	34850	G
17	35325	G
18	35631	C
19	35632	T
20	35721	C
21	35774	C
22	35992	C
23	36054	A
24	36207	C
25	36330	G
26	36617	C

(a) PS = polymorphic site;
(b) Position of PS within SEQ ID NO:1;
(c) Alleles for haplotypes are presented 5´ to 3´ in each column;

and wherein the haplotype pair is selected from the haplotype pairs shown in the table immediately below, wherein each of the POR haplotype pairs consists of first and second haplotypes which comprise first and second sequences of polymorphisms whose positions in SEQ ID NO:1 and identities are set forth in the table immediately below:

115

PS No.(a)	PS Position(b)	Haplotype Pair(c) (Part 1)							
		8/8	10/10	19/19	15/15	8/11	8/3	19/21	10/17
1	22635	G/G	G/G	G/G	G/G	G/G	G/A	G/G	G/G
2	29577	C/C	C/C	C/C	C/C	C/C	C/C	C/T	C/C
3	29754	C/C	C/C	T/T	C/C	C/C	C/C	T/C	C/C
4	30427	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
5	31531	A/A	A/A	A/A	G/G	A/A	A/A	A/A	A/G
6	31644	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
7	31648	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
8	32446	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
9	32478	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
10	33538	G/G	G/G	G/G	C/C	G/G	G/G	G/G	G/C
11	33551	A/A	A/A	A/A	G/G	A/A	A/A	A/A	A/G
12	33571	C/C	T/T	C/C	C/C	C/T	C/T	C/C	T/C
13	34766	T/T	C/C	T/T	C/C	T/T	T/C	T/T	C/C
14	34796	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/T
15	34797	C/C	C/C	C/C	T/T	C/C	C/C	C/C	C/T
16	34850	G/G	G/G	G/G	C/C	G/G	G/G	G/G	G/C
17	35325	G/G	G/G	G/G	G/G	G/G	G/A	G/G	G/G
18	35631	C/C	C/C	C/C	T/T	C/C	C/T	C/C	C/T
19	35632	T/T	T/T	T/T	G/G	T/T	T/T	T/T	T/G
20	35721	C/C	C/C	C/C	T/T	C/C	C/C	C/C	C/T
21	35774	C/C	T/T	C/C	C/C	C/T	C/T	C/C	T/C
22	35992	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
23	36054	G/G	G/G	A/A	G/G	G/G	G/G	A/A	G/G
24	36207	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
25	36330	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
26	36617	C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/C

145

(a) PS = polymorphic site;
(b) Position of PS in SEQ ID NO:1;
(c) Haplotype pairs are represented as 1st haplotype/2nd haplotype; with alleles of each haplotype shown 5′ to 3′ as 1st polymorphism/2nd polymorphism in each column;

150

150

PS No.(a)	PS Position(b)	Haplotype Pair(c) (Part 2)							
		15/2	19/14	8/16	8/20	19/2	19/4	14/9	8/13
1	22635	G/A	G/G	G/G	G/G	G/A	G/A	G/G	G/G
2	29577	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
3	29754	C/C	T/C	C/C	C/T	T/C	T/T	C/C	C/C
4	30427	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
5	31531	G/A	A/G	A/G	A/A	A/A	A/A	G/A	A/A
6	31644	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
7	31648	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/T
8	32446	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
9	32478	C/C	C/C	C/C	C/T	C/C	C/C	C/C	C/C
10	33538	C/G	G/C	G/C	G/G	G/G	G/G	C/G	G/G
11	33551	G/A	A/G	A/G	A/A	A/A	A/A	G/A	A/A
12	33571	C/T	C/C	C/C	C/C	C/T	C/C	C/T	C/T
13	34766	C/C	T/C	T/C	T/T	T/C	T/T	C/C	T/C
14	34796	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
15	34797	T/C	C/T	C/T	C/C	C/C	C/C	T/C	C/C
16	34850	C/G	G/C	G/G	G/G	G/G	G/G	C/G	G/G
17	35325	G/A	G/G	G/G	G/G	G/A	G/G	G/G	G/G
18	35631	T/C	C/T	C/T	C/C	C/C	C/C	T/C	C/C
19	35632	G/T	T/G	T/G	T/T	T/T	T/T	G/T	T/T
20	35721	T/C	C/C	C/T	C/C	C/C	C/C	C/C	C/C
21	35774	C/T	C/C	C/C	C/C	C/T	C/C	C/T	C/T
22	35992	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
23	36054	G/G	A/G	G/G	G/A	A/G	A/A	G/A	G/G
24	36207	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
25	36330	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
26	36617	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C

155
160
165
170
175

(a) PS = polymorphic site;

180

(b) Position of PS in SEQ ID NO:1;

(c) Haplotype pairs are represented as 1st haplotype/2nd haplotype; with alleles of each haplotype shown 5′ to 3′ as 1st polymorphism/2nd polymorphism in each column;

	PS No.(a)	PS Position(b)	Haplotype Pair(c) (Part 3)							
			8/12	8/1	8/6	8/15	8/17	8/7	10/5	8/2
185	1	22635	G/G	G/A	G/G	G/G	G/G	G/G	G/G	G/A
	2	29577	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	3	29754	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	4	30427	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	5	31531	A/A	A/A	A/A	A/G	A/G	A/A	A/A	A/A
190	6	31644	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	7	31648	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	8	32446	A/G	A/A	A/A	A/A	A/A	A/A	A/A	A/A
	9	32478	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	10	33538	G/G	G/G	G/G	G/C	G/C	G/G	G/G	G/G
195	11	33551	A/A	A/A	A/A	A/G	A/G	A/A	A/A	A/A
	12	33571	C/T	C/C	C/C	C/C	C/C	C/C	T/C	C/T
	13	34766	T/C	T/T	T/C	T/C	T/C	T/T	C/C	T/C
	14	34796	C/C	C/C	C/C	C/C	C/T	C/C	C/C	C/C
	15	34797	C/C	C/C	C/C	C/T	C/T	C/C	C/C	C/C
200	16	34850	G/G	G/G	G/G	G/C	G/C	G/G	G/G	G/G
	17	35325	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/A
	18	35631	C/C	C/C	C/C	C/T	C/T	C/C	C/C	C/C
	19	35632	T/T	T/T	T/T	T/G	T/G	T/T	T/T	T/T
	20	35721	C/C	C/C	C/C	C/T	C/T	C/C	C/C	C/C
205	21	35774	C/T	C/C	C/T	C/C	C/C	C/C	T/C	C/T
	22	35992	C/C	C/C	C/T	C/C	C/C	C/C	C/C	C/C
	23	36054	G/G	G/G	G/G	G/G	G/G	G/G	G/A	G/G
	24	36207	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
	25	36330	G/G	G/G	G/G	G/G	G/G	G/A	G/G	G/G
210	26	36617	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C

(a) PS = polymorphic site;

(b) Position of PS in SEQ ID NO:1;

(c) Haplotype pairs are represented as 1st haplotype/2nd haplotype; with alleles of each

215

haplotype shown 5′ to 3′ as 1st polymorphism/2nd polymorphism in each column;

PS No.(a)	PS Position(b)	Haplotype Pair(c) (Part 4) 8/19	10/15	8/10	10/19	19/17	15/18	19/15
1	22635	G/G	G/G	G/G	G/G	G/G	G/G	G/G
2	29577	C/C	C/C	C/C	C/C	C/C	C/C	C/C
3	29754	C/T	C/C	C/C	C/T	T/C	C/C	T/C
4	30427	C/C	C/C	C/C	C/C	C/C	C/T	C/C
5	31531	A/A	A/G	A/A	A/A	A/G	G/G	A/G
6	31644	C/C	C/C	C/C	C/C	C/C	C/C	C/C
7	31648	C/C	C/C	C/C	C/C	C/C	C/C	C/C
8	32446	A/A	A/A	A/A	A/A	A/A	A/A	A/A
9	32478	C/C	C/C	C/C	C/C	C/C	C/C	C/C
10	33538	G/G	G/C	G/G	G/G	G/C	C/C	G/C
11	33551	A/A	A/G	A/A	A/A	A/G	G/G	A/G
12	33571	C/C	T/C	C/T	T/C	C/C	C/C	C/C
13	34766	T/T	C/C	T/C	C/T	T/C	C/C	T/C
14	34796	C/C	C/C	C/C	C/C	C/T	C/C	C/C
15	34797	C/C	C/T	C/C	C/C	C/T	T/T	C/T
16	34850	G/G	G/C	G/G	G/G	G/C	C/C	G/C
17	35325	G/G	G/G	G/G	G/G	G/G	G/G	G/G
18	35631	C/C	C/T	C/C	C/C	C/T	T/T	C/T
19	35632	T/T	T/G	T/T	T/T	T/G	G/G	T/G
20	35721	C/C	C/T	C/C	C/C	C/T	T/C	C/T
21	35774	C/C	T/C	C/T	T/C	C/C	C/C	C/C
22	35992	C/C	C/C	C/C	C/C	C/C	C/C	C/C
23	36054	G/A	G/G	G/G	G/A	A/G	G/G	A/G
24	36207	C/C	C/C	C/C	C/C	C/C	C/C	C/C
25	36330	G/G	G/G	G/G	G/G	G/G	G/G	G/G
26	36617	C/C	C/C	C/C	C/C	C/C	C/C	C/C

(a) PS = polymorphic site;

(b) Position of PS in SEQ ID NO:1;

(c) Haplotype pairs are represented as 1st haplotype/2nd haplotype; with alleles of each haplotype shown 5′ to 3′ as 1st polymorphism/2nd polymorphism in each column;

wherein a higher frequency of the haplotype or haplotype pair in the trait population than in the reference population indicates the trait is associated with the haplotype or haplotype pair.

11. The method of claim 10, wherein the trait is a clinical response to a drug targeting POR or to a drug for treating a condition or disease associated with POR activity.

12. An isolated oligonucleotide designed for detecting a polymorphism in the P450 (cytochrome) oxidoreductase (POR) gene at a polymorphic site (PS) selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18, PS19, PS20, PS21, PS22, PS23, PS24, PS25 and PS26, wherein the selected PS have the position and alternative alleles shown in SEQ ID NO:1.

13. The isolated oligonucleotide of claim 12, which is an allele-specific oligonucleotide that specifically hybridizes to an allele of the POR gene at a region containing the polymorphic site.

14. The allele-specific oligonucleotide of claim 13, which comprises a nucleotide sequence selected from the group consisting of SEQ ID NOS:4-29, the complements of SEQ ID NOS:4-29, and SEQ ID NOS:30-81.

15. The isolated oligonucleotide of claim 12, which is a primer-extension oligonucleotide.

16. The primer-extension oligonucleotide of claim 15,which comprises a nucleotide sequence selected from the group consisting of SEQ ID NOS:82-133.

17. A kit for haplotyping or genotyping the P450 (cytochrome) oxidoreductase (POR) gene of an individual, which comprises a set of oligonucleotides designed to haplotype or genotype each of polymorphic sites (PS) PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18, PS19, PS20, PS21, PS22, PS23, PS24, PS25 and PS26, wherein the selected PS have the position and alternative alleles shown in SEQ ID NO:1.

18. An isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of:

   (a) a first nucleotide sequence which comprises a P450 (cytochrome) oxidoreductase (POR) isogene, wherein the POR isogene is selected from the group consisting of isogenes 1-21 shown in the table immediately below and wherein each of the isogenes comprises the regions of SEQ ID NO:1 shown in the table immediately below and wherein each of the isogenes 1-21 is further defined by the corresponding sequence of polymorphisms whose positions and identities are set forth in the table immediately below; and

Region Examined(a)	PS No.(b)	PS Position(c)	Isogene Number(d) (Part 1)									
			1	2	3	4	5	6	7	8	9	10
3878-4429	-	-	-	-	-	-	-	-	-	-	-	-
22306-22748	1	22635	A	A	A	A	G	G	G	G	G	G
29363-29820	2	29577	C	C	C	C	C	C	C	C	C	C
29363-29820	3	29754	C	C	C	T	C	C	C	C	C	C
30215-30747	4	30427	C	C	C	C	C	C	C	C	C	C
30942-31856	5	31531	A	A	A	A	A	A	A	A	A	A
30942-31856	6	31644	C	C	C	C	C	C	C	C	C	C
30942-31856	7	31648	C	C	C	C	C	C	C	C	C	C
32039-32600	8	32446	A	A	A	A	A	A	A	A	A	A
32039-32600	9	32478	C	C	C	C	C	C	C	C	C	C
33448-34167	10	33538	G	G	G	G	G	G	G	G	G	G
33448-34167	11	33551	A	A	A	A	A	A	A	A	A	A
33448-34167	12	33571	C	T	T	C	C	C	C	C	T	T
34661-36857	13	34766	T	C	C	T	C	C	T	T	C	C
34661-36857	14	34796	C	C	C	C	C	C	C	C	C	C
34661-36857	15	34797	C	C	C	C	C	C	C	C	C	C
34661-36857	16	34850	G	G	G	G	G	G	G	G	G	G
34661-36857	17	35325	G	A	A	G	G	G	G	G	G	G
34661-36857	18	35631	C	C	T	C	C	C	C	C	C	C
34661-36857	19	35632	T	T	T	T	T	T	T	T	T	T
34661-36857	20	35721	C	C	C	C	C	C	C	C	C	C
34661-36857	21	35774	C	T	T	C	C	T	C	C	T	T
34661-36857	22	35992	C	C	C	C	C	T	C	C	C	C
34661-36857	23	36054	G	G	G	A	A	G	G	G	A	G
34661-36857	24	36207	C	C	C	C	C	C	C	C	C	C
34661-36857	25	36330	G	G	G	G	G	G	A	G	G	G
34661-36857	26	36617	C	C	T	C	C	C	C	C	C	C

Region Examined(a)	PS No.(b)	PS Position(c)	Isogene Number(d) (Part 2)									
			11	12	13	14	15	16	17	18	19	20
3878-4429	-	-	-	-	-	-	-	-	-	-	-	-
22306-22748	1	22635	G	G	G	G	G	G	G	G	G	G
29363-29820	2	29577	C	C	C	C	C	C	C	C	C	C
29363-29820	3	29754	C	C	C	C	C	C	C	C	T	T
30215-30747	4	30427	C	C	C	C	C	C	C	T	C	C
30942-31856	5	31531	A	A	A	G	G	G	G	G	A	A
30942-31856	6	31644	C	C	C	C	C	C	C	C	C	C
30942-31856	7	31648	C	C	T	C	C	C	C	C	C	C
32039-32600	8	32446	A	G	A	A	A	A	A	A	A	A
32039-32600	9	32478	C	C	C	C	C	C	C	C	C	T
33448-34167	10	33538	G	G	G	C	C	C	C	C	G	G
33448-34167	11	33551	A	A	A	G	G	G	G	G	A	A
33448-34167	12	33571	T	T	T	C	C	C	C	C	C	C
34661-36857	13	34766	T	C	C	C	C	C	C	C	T	T
34661-36857	14	34796	C	C	C	C	C	C	T	C	C	C
34661-36857	15	34797	C	C	C	T	T	T	T	T	C	C
34661-36857	16	34850	G	G	G	C	C	G	C	C	G	G
34661-36857	17	35325	G	G	G	G	G	G	G	G	G	G
34661-36857	18	35631	C	C	C	T	T	T	T	T	C	C
34661-36857	19	35632	T	T	T	G	G	G	G	G	T	T
34661-36857	20	35721	C	C	C	C	T	T	T	C	C	C
34661-36857	21	35774	T	T	T	C	C	C	C	C	C	C
34661-36857	22	35992	C	C	C	C	C	C	C	C	C	C
34661-36857	23	36054	G	G	G	G	G	G	G	G	A	A
34661-36857	24	36207	C	C	C	C	C	C	C	C	C	C
34661-36857	25	36330	G	G	G	G	G	G	G	G	G	G
34661-36857	26	36617	C	C	C	C	C	C	C	C	C	C

(a) Region examined represents the nucleotide positions defining the start and stop positions within the 1ˢᵗ SEQ ID NO of the sequenced region.

(b) PS = polymorphic site;

(c) Position of PS in SEQ ID NO:1;

(d) Alleles for isogenes are presented 5′ to 3′ in each column;

Region Examined(a)	PS No.(b)	PS Position(c)	Isogene Number(d) (Part 3) 21
3878-4429	-	-	-
22306-22748	1	22635	G
29363-29820	2	29577	T
29363-29820	3	29754	C
30215-30747	4	30427	C
30942-31856	5	31531	A
30942-31856	6	31644	C
30942-31856	7	31648	C
32039-32600	8	32446	A
32039-32600	9	32478	C
33448-34167	10	33538	G
33448-34167	11	33551	A
33448-34167	12	33571	C
34661-36857	13	34766	T
34661-36857	14	34796	C
34661-36857	15	34797	C
34661-36857	16	34850	G
34661-36857	17	35325	G
34661-36857	18	35631	C
34661-36857	19	35632	T
34661-36857	20	35721	C
34661-36857	21	35774	C
34661-36857	22	35992	C
34661-36857	23	36054	A
34661-36857	24	36207	C
34661-36857	25	36330	G
34661-36857	26	36617	C

(a) Region examined represents the nucleotide positions defining the start and stop positions within the 1st SEQ ID NO of the sequenced region.

(b) PS = polymorphic site;

(c) Position of PS in SEQ ID NO:1;

(d) Alleles for isogenes are presented 5′ to 3′ in each column;

(b) a second nucleotide sequence which is complementary to the first nucleotide sequence.

19. The isolated polynucleotide of claim 18, which is a DNA molecule and comprises both the first and second nucleotide sequences and further comprises expression regulatory elements operably linked to the first nucleotide sequence.

20. A recombinant nonhuman organism transformed or transfected with the isolated polynucleotide of claim 19, wherein the organism expresses a POR protein that is encoded by the first nucleotide sequence.

21. The recombinant nonhuman organism of claim 20, which is a transgenic animal.

22. An isolated fragment of a P450 (cytochrome) oxidoreductase (POR) isogene, wherein the fragment comprises at least 10 nucleotides in one of the regions of SEQ ID NO:1 shown in the table immediately below and wherein the fragment comprises one or more polymorphisms selected from the group consisting of adenine at PS1, thymine at PS2, thymine at PS3, thymine at PS4, adenine at PS5, thymine at PS6, thymine at PS7, guanine at PS8, thymine at PS9,

5

68

guanine at PS10, adenine at PS11, thymine at PS12, thymine at PS13, thymine at PS14, cytosine at PS15, guanine at PS16, adenine at PS17, cytosine at PS18, thymine at PS19, thymine at PS20, cytosine at PS21, thymine at PS22, guanine at PS23, guanine at PS24, adenine at PS25 and cytosine at PS26, wherein the selected polymorphism has the position set forth in the table immediately below:

Region Examined(a)	PS No.(b)	PS Position(c)	Isogene Number(d) (Part 1)									
			1	2	3	4	5	6	7	8	9	10
3878-4429	-	-	-	-	-	-	-	-	-	-	-	-
22306-22748	1	22635	A	A	A	A	G	G	G	G	G	G
29363-29820	2	29577	C	C	C	C	C	C	C	C	C	C
29363-29820	3	29754	C	C	C	T	C	C	C	C	C	C
30215-30747	4	30427	C	C	C	C	C	C	C	C	C	C
30942-31856	5	31531	A	A	A	A	A	A	A	A	A	A
30942-31856	6	31644	C	C	C	C	C	C	C	C	C	C
30942-31856	7	31648	C	C	C	C	C	C	C	C	C	C
32039-32600	8	32446	A	A	A	A	A	A	A	A	A	A
32039-32600	9	32478	C	C	C	C	C	C	C	C	C	C
33448-34167	10	33538	G	G	G	G	G	G	G	G	G	G
33448-34167	11	33551	A	A	A	A	A	A	A	A	A	A
33448-34167	12	33571	C	T	T	C	C	C	C	C	T	T
34661-36857	13	34766	T	C	C	T	C	C	T	T	C	C
34661-36857	14	34796	C	C	C	C	C	C	C	C	C	C
34661-36857	15	34797	C	C	C	C	C	C	C	C	C	C
34661-36857	16	34850	G	G	G	G	G	G	G	G	G	G
34661-36857	17	35325	G	A	A	G	G	G	G	G	G	G
34661-36857	18	35631	C	C	T	C	C	C	C	C	C	C
34661-36857	19	35632	T	T	T	T	T	T	T	T	T	T
34661-36857	20	35721	C	C	C	C	C	C	C	C	C	C
34661-36857	21	35774	C	T	T	C	C	T	C	C	T	T
34661-36857	22	35992	C	C	C	C	C	T	C	C	C	C
34661-36857	23	36054	G	G	G	A	A	G	G	G	A	G
34661-36857	24	36207	C	C	C	C	C	C	C	C	C	C
34661-36857	25	36330	G	G	G	G	G	G	A	G	G	G
34661-36857	26	36617	C	C	T	C	C	C	C	C	C	C

(a) Region examined represents the nucleotide positions defining the start and stop positions within SEQ ID NO:1 of the regions sequenced;
(b) PS = polymorphic site;
(c) Position of PS within SEQ ID NO:1;
(d) Alleles for POR isogenes are presented 5' to 3' in each column;

Region Examined(a)	PS No.(b)	PS Position(c)	Isogene Number(d) (Part 2)									
			11	12	13	14	15	16	17	18	19	20
3878-4429	-	-	-	-	-	-	-	-	-	-	-	-
22306-22748	1	22635	G	G	G	G	G	G	G	G	G	G
29363-29820	2	29577	C	C	C	C	C	C	C	C	C	C
29363-29820	3	29754	C	C	C	C	C	C	C	C	T	T
30215-30747	4	30427	C	C	C	C	C	C	C	T	C	C
30942-31856	5	31531	A	A	A	G	G	G	G	G	A	A
30942-31856	6	31644	C	C	C	C	C	C	C	C	C	C
30942-31856	7	31648	C	C	T	C	C	C	C	C	C	C
32039-32600	8	32446	A	G	A	A	A	A	A	A	A	A
32039-32600	9	32478	C	C	C	C	C	C	C	C	C	T
33448-34167	10	33538	G	G	G	C	C	C	C	C	G	G
33448-34167	11	33551	A	A	A	G	G	G	G	G	A	A
33448-34167	12	33571	T	T	T	C	C	C	C	C	C	C
34661-36857	13	34766	T	C	C	C	C	C	C	C	T	T
34661-36857	14	34796	C	C	C	C	C	C	T	C	C	C
34661-36857	15	34797	C	C	C	T	T	T	T	T	C	C
34661-36857	16	34850	G	G	G	C	C	G	C	C	G	G
34661-36857	17	35325	G	G	G	G	G	G	G	G	G	G
34661-36857	18	35631	C	C	C	T	T	T	T	T	C	C
34661-36857	19	35632	T	T	T	G	G	G	G	G	T	T
34661-36857	20	35721	C	C	C	C	T	T	T	C	C	C
34661-36857	21	35774	T	T	T	C	C	C	C	C	C	C
34661-36857	22	35992	C	C	C	C	C	C	C	C	C	C
34661-36857	23	36054	G	G	G	G	G	G	G	G	A	A
34661-36857	24	36207	C	C	C	C	C	C	C	C	C	C
34661-36857	25	36330	G	G	G	G	G	G	G	G	G	G
34661-36857	26	36617	C	C	C	C	C	C	C	C	C	C

(a) Region examined represents the nucleotide positions defining the start and stop positions within SEQ ID NO:1 of the regions sequenced;

(b) PS = polymorphic site;

(c) Position of PS within SEQ ID NO:1;

(d) Alleles for POR isogenes are presented 5′ to 3′ in each column;

Region Examined(a)	PS No.(b)	PS Position(c)	Isogene Number(d) (Part 3) 21
3878-4429	-	-	-
22306-22748	1	22635	G
29363-29820	2	29577	T
29363-29820	3	29754	C
30215-30747	4	30427	C
30942-31856	5	31531	A
30942-31856	6	31644	C
30942-31856	7	31648	C
32039-32600	8	32446	A
32039-32600	9	32478	C
33448-34167	10	33538	G
33448-34167	11	33551	A
33448-34167	12	33571	C
34661-36857	13	34766	T
34661-36857	14	34796	C
34661-36857	15	34797	C
34661-36857	16	34850	G
34661-36857	17	35325	G
34661-36857	18	35631	C
34661-36857	19	35632	T
34661-36857	20	35721	C
34661-36857	21	35774	C
34661-36857	22	35992	C
34661-36857	23	36054	A
34661-36857	24	36207	C
34661-36857	25	36330	G
34661-36857	26	36617	C

(a) Region examined represents the nucleotide positions defining the start and stop positions within SEQ ID NO:1 of the regions sequenced;

(b) PS = polymorphic site;

(c) Position of PS within SEQ ID NO:1;

(d) Alleles for POR isogenes are presented 5' to 3' in each column.

23. An isolated polynucleotide comprising a POR coding sequence, wherein the coding sequence is selected from the group consisting of 1-21 shown in the table immediately below, and wherein each of the coding sequences comprises the regions of SEQ ID NO:2 that are defined by exons 2-16, except at each of the polymorphic sites which have the positions in SEQ ID NO:2 and polymorphisms set forth in the table immediately below:

PS No.(a)	PS Position(b)	Isogene Coding Sequence Number(c) (Part 1)									
		1c	2c	3c	4c	5c	6c	7c	8c	9c	10c
2	266	C	C	C	C	C	C	C	C	C	C
4	357	C	C	C	C	C	C	C	C	C	C
6	671	C	C	C	C	C	C	C	C	C	C
7	675	C	C	C	C	C	C	C	C	C	C
20	1443	C	C	C	C	C	C	C	C	C	C
21	1496	C	T	T	C	C	T	C	C	T	T
23	1703	G	G	G	A	A	G	G	G	A	G

PS No.(a)	PS Position(b)	Isogene Coding Sequence Number(c) (Part 2)									
		11c	12c	13c	14c	15c	16c	17c	18c	19c	20c
2	266	C	C	C	C	C	C	C	C	C	C
4	357	C	C	C	C	C	C	C	T	C	C
6	671	C	C	C	C	C	C	C	C	C	C
7	675	C	C	T	C	C	C	C	C	C	C
20	1443	C	C	C	C	T	T	T	C	C	C
21	1496	T	T	T	C	C	C	C	C	C	C
23	1703	G	G	G	G	G	G	G	G	A	A

PS No.(a)	PS Position(b)	Isogene Coding Sequence Number(c) (Part 3)
		21c
2	266	T
4	357	C
6	671	C
7	675	C
20	1443	C
21	1496	C
23	1703	A

(a) PS = polymorphic site;

(b) Position of PS in SEQ ID NO:2;

(c) Alleles for the isogene coding sequence are presented 5′ to 3′ in each column; the numerical portion of the isogene coding sequence number represents the number of the parent full POR isogene.

24. A recombinant nonhuman organism transformed or transfected with the isolated polynucleotide of claim 23, wherein the organism expresses a P450 (cytochrome) oxidoreductase (POR) protein that is encoded by the polymorphic variant sequence.

25. The recombinant nonhuman organism of claim 24, which is a transgenic animal.

26. An isolated fragment of a POR cDNA, wherein the fragment comprises one or more polymorphisms selected from the group consisting of thymine at a position corresponding to nucleotide 266, thymine at a position corresponding to nucleotide 357, thymine at a position corresponding to nucleotide 671, thymine at a position corresponding to nucleotide 675, thymine at a position corresponding to nucleotide 1443, cytosine at a position corresponding to nucleotide 1496 and guanine at a position corresponding to nucleotide 1703 in SEQ ID NO:2.

27 An isolated polypeptide comprising an amino acid sequence which is a polymorphic variant of a reference sequence for the P450 (cytochrome) oxidoreductase (POR) protein, wherein the

reference sequence comprises SEQ ID NO:3 for the regions encoded by exons 2-16, except the polymorphic variant comprises one or more variant amino acids selected from the group consisting of isoleucine at a position corresponding to amino acid position 89, leucine at a position corresponding to amino acid position 224, alanine at a position corresponding to amino acid position 499 and tryptophan at a position corresponding to amino acid position 568.

28. . An isolated monoclonal antibody specific for and immunoreactive with the isolated polypeptide of claim 27.

29. A method for screening for drugs targeting the isolated polypeptide of claim 27 which comprises contacting the POR polymorphic variant with a candidate agent and assaying for binding activity.

30. An isolated fragment of the POR polypeptide, wherein the fragment comprises one or more variant amino acids selected from the group consisting of isoleucine at a position corresponding to amino acid position 89, leucine at a position corresponding to amino acid position 224, alanine at a position corresponding to amino acid position 499 and tryptophan at a position corresponding to amino acid position 568 in SEQ ID NO:3.

31 ' A computer system for storing and analyzing polymorphism data for the P450 (cytochrome) oxidoreductase gene, comprising:

    (a)   a central processing unit (CPU);

    (b)   a communication interface;

    (c)   a display device;

5   (d)   an input device; and

    (e)   a database containing the polymorphism data;

wherein the polymorphism data comprises any one or more of the haplotypes set forth in the table immediately below:

10

PS No.(a)	PS Position(b)	Haplotype Number(c) (Part 1)									
		1	2	3	4	5	6	7	8	9	10
1	22635	A	A	A	A	G	G	G	G	G	G
2	29577	C	C	C	C	C	C	C	C	C	C
3	29754	C	C	C	T	C	C	C	C	C	C
4	30427	C	C	C	C	C	C	C	C	C	C
5	31531	A	A	A	A	A	A	A	A	A	A
6	31644	C	C	C	C	C	C	C	C	C	C
7	31648	C	C	C	C	C	C	C	C	C	C
8	32446	A	A	A	A	A	A	A	A	A	A
9	32478	C	C	C	C	C	C	C	C	C	C
10	33538	G	G	G	G	G	G	G	G	G	G
11	33551	A	A	A	A	A	A	A	A	A	A
12	33571	C	T	T	C	C	C	C	C	T	T
13	34766	T	C	C	T	C	C	T	T	C	C
14	34796	C	C	C	C	C	C	C	C	C	C
15	34797	C	C	C	C	C	C	C	C	C	C
16	34850	G	G	G	G	G	G	G	G	G	G
17	35325	G	A	A	G	G	G	G	G	G	G
18	35631	C	C	T	C	C	C	C	C	C	C
19	35632	T	T	T	T	T	T	T	T	T	T
20	35721	C	C	C	C	C	C	C	C	C	C
21	35774	C	T	T	C	C	T	C	C	T	T
22	35992	C	C	C	C	C	T	C	C	C	C
23	36054	G	G	G	A	A	G	G	G	A	G
24	36207	C	C	C	C	C	C	C	C	C	C
25	36330	G	G	G	G	G	G	A	G	G	G
26	36617	C	C	T	C	C	C	C	C	C	C

(a) PS = polymorphic site;
(b) Position of PS within SEQ ID NO:1;
(c) Alleles for haplotypes are presented 5´ to 3´ in each column;

PS No.(a)	PS Position(b)	Haplotype Number(c) (Part 2)									
		11	12	13	14	15	16	17	18	19	20
1	22635	G	G	G	G	G	G	G	G	G	G
2	29577	C	C	C	C	C	C	C	C	C	C
3	29754	C	C	C	C	C	C	C	C	T	T
4	30427	C	C	C	C	C	C	C	T	C	C
5	31531	A	A	A	G	G	G	G	G	A	A
6	31644	C	C	C	C	C	C	C	C	C	C
7	31648	C	C	T	C	C	C	C	C	C	C
8	32446	A	G	A	A	A	A	A	A	A	A
9	32478	C	C	C	C	C	C	C	C	C	T
10	33538	G	G	G	C	C	C	C	C	G	G
11	33551	A	A	A	G	G	G	G	G	A	A
12	33571	T	T	T	C	C	C	C	C	C	C
13	34766	T	C	C	C	C	C	C	C	T	T
14	34796	C	C	C	C	C	C	T	C	C	C
15	34797	C	C	C	T	T	T	T	T	C	C
16	34850	G	G	G	C	C	G	C	C	G	G
17	35325	G	G	G	G	G	G	G	G	G	G
18	35631	C	C	C	T	T	T	T	T	C	C
19	35632	T	T	T	G	G	G	G	G	T	T
20	35721	C	C	C	C	T	T	T	C	C	C
21	35774	T	T	T	C	C	C	C	C	C	C
22	35992	C	C	C	C	C	C	C	C	C	C
23	36054	G	G	G	G	G	G	G	G	A	A
24	36207	C	C	C	C	C	C	C	C	C	C
25	36330	G	G	G	G	G	G	G	G	G	G
26	36617	C	C	C	C	C	C	C	C	C	C

(a) PS = polymorphic site;
(b) Position of PS within SEQ ID NO:1;
(c) Alleles for haplotypes are presented 5′ to 3′ in each column;

PS No.(a)	PS Position(b)	Haplotype Number(c) (Part 3)
		21
1	22635	G
2	29577	T
3	29754	C
4	30427	C
5	31531	A
6	31644	C
7	31648	C
8	32446	A
9	32478	C
10	33538	G
11	33551	A
12	33571	C
13	34766	T
14	34796	C
15	34797	C
16	34850	G
17	35325	G
18	35631	C
19	35632	T
20	35721	C
21	35774	C
22	35992	C
23	36054	A
24	36207	C
25	36330	G
26	36617	C

80, 85, 90, 95, 100

(a) PS = polymorphic site;
(b) Position of PS within SEQ ID NO:1;
(c) Alleles for haplotypes are presented 5′ to 3′ in each column;

the haplotype pairs set forth in the table immediately below:

105

110

PS No.(a)	PS Position(b)	Haplotype Pair(c) (Part 1)							
		8/8	10/10	19/19	15/15	8/11	8/3	19/21	10/17
1	22635	G/G	G/G	G/G	G/G	G/G	G/A	G/G	G/G
2	29577	C/C	C/C	C/C	C/C	C/C	C/C	C/T	C/C
3	29754	C/C	C/C	T/T	C/C	C/C	C/C	T/C	C/C
4	30427	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
5	31531	A/A	A/A	A/A	G/G	A/A	A/A	A/A	A/G
6	31644	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
7	31648	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
8	32446	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
9	32478	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
10	33538	G/G	G/G	G/G	C/C	G/G	G/G	G/G	G/C
11	33551	A/A	A/A	A/A	G/G	A/A	A/A	A/A	A/G
12	33571	C/C	T/T	C/C	C/C	C/T	C/T	C/C	T/C
13	34766	T/T	C/C	T/T	C/C	T/T	T/C	T/T	C/C
14	34796	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/T
15	34797	C/C	C/C	C/C	T/T	C/C	C/C	C/C	C/T
16	34850	G/G	G/G	G/G	C/C	G/G	G/G	G/G	G/C
17	35325	G/G	G/G	G/G	G/G	G/G	G/A	G/G	G/G
18	35631	C/C	C/C	C/C	T/T	C/C	C/T	C/C	C/T
19	35632	T/T	T/T	T/T	G/G	T/T	T/T	T/T	T/G
20	35721	C/C	C/C	C/C	T/T	C/C	C/C	C/C	C/T
21	35774	C/C	T/T	C/C	C/C	C/T	C/T	C/C	T/C
22	35992	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
23	36054	G/G	G/G	A/A	G/G	G/G	G/G	A/A	G/G
24	36207	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
25	36330	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
26	36617	C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/C

(a) PS = polymorphic site;
(b) Position of PS in SEQ ID NO:1;
(c) Haplotype pairs are represented as 1st haplotype/2nd haplotype; with alleles of each haplotype shown 5′ to 3′ as 1st polymorphism/2nd polymorphism in each column;

PS No.(a)	PS Position(b)	Haplotype Pair(c) (Part 2)							
		15/2	19/14	8/16	8/20	19/2	19/4	14/9	8/13
1	22635	G/A	G/G	G/G	G/G	G/A	G/A	G/G	G/G
2	29577	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
3	29754	C/C	T/C	C/C	C/T	T/C	T/T	C/C	C/C
4	30427	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
5	31531	G/A	A/G	A/G	A/A	A/A	A/A	G/A	A/A
6	31644	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
7	31648	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/T
8	32446	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
9	32478	C/C	C/C	C/C	C/T	C/C	C/C	C/C	C/C
10	33538	C/G	G/C	G/C	G/G	G/G	G/G	C/G	G/G
11	33551	G/A	A/G	A/G	A/A	A/A	A/A	G/A	A/A
12	33571	C/T	C/C	C/C	C/C	C/T	C/C	C/T	C/T
13	34766	C/C	T/C	T/C	T/T	T/C	T/T	C/C	T/C
14	34796	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
15	34797	T/C	C/T	C/T	C/C	C/C	C/C	T/C	C/C
16	34850	C/G	G/C	G/G	G/G	G/G	G/G	C/G	G/G
17	35325	G/A	G/G	G/G	G/G	G/A	G/G	G/G	G/G
18	35631	T/C	C/T	C/T	C/C	C/C	C/C	T/C	C/C
19	35632	G/T	T/G	T/G	T/T	T/T	T/T	G/T	T/T
20	35721	T/C	C/C	C/T	C/C	C/C	C/C	C/C	C/C
21	35774	C/T	C/C	C/C	C/C	C/T	C/C	C/T	C/T
22	35992	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
23	36054	G/G	A/G	G/G	G/A	A/G	A/A	G/A	G/G
24	36207	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
25	36330	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
26	36617	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C

(a) PS = polymorphic site;

(b) Position of PS in SEQ ID NO:1;

(c) Haplotype pairs are represented as 1st haplotype/2nd haplotype; with alleles of each haplotype shown 5' to 3' as 1st polymorphism/2nd polymorphism in each column;

PS No.(a)	PS Position(b)	Haplotype Pair(c) (Part 3)							
		8/12	8/1	8/6	8/15	8/17	8/7	10/5	8/2
1	22635	G/G	G/A	G/G	G/G	G/G	G/G	G/G	G/A
2	29577	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
3	29754	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
4	30427	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
5	31531	A/A	A/A	A/A	A/G	A/G	A/A	A/A	A/A
6	31644	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
7	31648	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
8	32446	A/G	A/A	A/A	A/A	A/A	A/A	A/A	A/A
9	32478	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
10	33538	G/G	G/G	G/G	G/C	G/C	G/G	G/G	G/G
11	33551	A/A	A/A	A/A	A/G	A/G	A/A	A/A	A/A
12	33571	C/T	C/C	C/C	C/C	C/C	C/C	T/C	C/T
13	34766	T/C	T/T	T/C	T/C	T/C	T/T	C/C	T/C
14	34796	C/C	C/C	C/C	C/C	C/T	C/C	C/C	C/C
15	34797	C/C	C/C	C/C	C/T	C/T	C/C	C/C	C/C
16	34850	G/G	G/G	G/G	G/C	G/C	G/G	G/G	G/G
17	35325	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/A
18	35631	C/C	C/C	C/C	C/T	C/T	C/C	C/C	C/C
19	35632	T/T	T/T	T/T	T/G	T/G	T/T	T/T	T/T
20	35721	C/C	C/C	C/C	C/T	C/T	C/C	C/C	C/C
21	35774	C/T	C/C	C/T	C/C	C/C	C/C	T/C	C/T
22	35992	C/C	C/C	C/T	C/C	C/C	C/C	C/C	C/C
23	36054	G/G	G/G	G/G	G/G	G/G	G/G	G/A	G/G
24	36207	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
25	36330	G/G	G/G	G/G	G/G	G/G	G/A	G/G	G/G
26	36617	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C

(a) PS = polymorphic site;

(b) Position of PS in SEQ ID NO:1;

(c) Haplotype pairs are represented as 1st haplotype/2nd haplotype; with alleles of each haplotype shown 5′ to 3′ as 1st polymorphism/2nd polymorphism in each column;

PS No.(a)	PS Position(b)	Haplotype Pair(c) (Part 4)						
		8/19	10/15	8/10	10/19	19/17	15/18	19/15
1	22635	G/G	G/G	G/G	G/G	G/G	G/G	G/G
2	29577	C/C	C/C	C/C	C/C	C/C	C/C	C/C
3	29754	C/T	C/C	C/C	C/T	T/C	C/C	T/C
4	30427	C/C	C/C	C/C	C/C	C/C	C/T	C/C
5	31531	A/A	A/G	A/A	A/A	A/G	G/G	A/G
6	31644	C/C	C/C	C/C	C/C	C/C	C/C	C/C
7	31648	C/C	C/C	C/C	C/C	C/C	C/C	C/C
8	32446	A/A	A/A	A/A	A/A	A/A	A/A	A/A
9	32478	C/C	C/C	C/C	C/C	C/C	C/C	C/C
10	33538	G/G	G/C	G/G	G/G	G/C	C/C	G/C
11	33551	A/A	A/G	A/A	A/A	A/G	G/G	A/G
12	33571	C/C	T/C	C/T	T/C	C/C	C/C	C/C
13	34766	T/T	C/C	T/C	C/T	T/C	C/C	T/C
14	34796	C/C	C/C	C/C	C/C	C/T	C/C	C/C
15	34797	C/C	C/T	C/C	C/C	C/T	T/T	C/T
16	34850	G/G	G/C	G/G	G/G	G/C	C/C	G/C
17	35325	G/G	G/G	G/G	G/G	G/G	G/G	G/G
18	35631	C/C	C/T	C/C	C/C	C/T	T/T	C/T
19	35632	T/T	T/G	T/T	T/T	T/G	G/G	T/G
20	35721	C/C	C/T	C/C	C/C	C/T	T/C	C/T
21	35774	C/C	T/C	C/T	T/C	C/C	C/C	C/C
22	35992	C/C	C/C	C/C	C/C	C/C	C/C	C/C
23	36054	G/A	G/G	G/G	G/A	A/G	G/G	A/G
24	36207	C/C	C/C	C/C	C/C	C/C	C/C	C/C
25	36330	G/G	G/G	G/G	G/G	G/G	G/G	G/G
26	36617	C/C	C/C	C/C	C/C	C/C	C/C	C/C

(a) PS = polymorphic site;

(b) Position of PS in SEQ ID NO:1;

(c) Haplotype pairs are represented as 1st haplotype/2nd haplotype; with alleles of each haplotype shown 5′ to 3′ as 1st polymorphism/2nd polymorphism in each column;

and the frequency data in Tables 6 and 7.

32. A genome anthology for the P450 (cytochrome) oxidoreductase (POR) gene which comprises two or more POR isogenes selected from the group consisting of isogenes 1-21 shown in the table immediately below, and wherein each of the isogenes comprises the regions of SEQ ID NO:1 shown in the table immediately below and wherein each of the isogenes 1-21 is further defined by the corresponding sequence of polymorphisms whose positions and identities are set forth in the table immediately below:

Region Examined(a)	PS No.(b)	PS Position(c)	Isogene Number(d) (Part 1)									
			1	2	3	4	5	6	7	8	9	10
3878-4429	-	-	-	-	-	-	-	-	-	-	-	-
22306-22748	1	22635	A	A	A	A	G	G	G	G	G	G
29363-29820	2	29577	C	C	C	C	C	C	C	C	C	C
29363-29820	3	29754	C	C	C	T	C	C	C	C	C	C
30215-30747	4	30427	C	C	C	C	C	C	C	C	C	C
30942-31856	5	31531	A	A	A	A	A	A	A	A	A	A
30942-31856	6	31644	C	C	C	C	C	C	C	C	C	C
30942-31856	7	31648	C	C	C	C	C	C	C	C	C	C
32039-32600	8	32446	A	A	A	A	A	A	A	A	A	A
32039-32600	9	32478	C	C	C	C	C	C	C	C	C	C
33448-34167	10	33538	G	G	G	G	G	G	G	G	G	G
33448-34167	11	33551	A	A	A	A	A	A	A	A	A	A
33448-34167	12	33571	C	T	T	C	C	C	C	C	T	T
34661-36857	13	34766	T	C	C	T	C	C	T	T	C	C
34661-36857	14	34796	C	C	C	C	C	C	C	C	C	C
34661-36857	15	34797	C	C	C	C	C	C	C	C	C	C
34661-36857	16	34850	G	G	G	G	G	G	G	G	G	G
34661-36857	17	35325	G	A	A	G	G	G	G	G	G	G
34661-36857	18	35631	C	C	T	C	C	C	C	C	C	C
34661-36857	19	35632	T	T	T	T	T	T	T	T	T	T
34661-36857	20	35721	C	C	C	C	C	C	C	C	C	C
34661-36857	21	35774	C	T	T	C	C	T	C	C	T	T
34661-36857	22	35992	C	C	C	C	C	T	C	C	C	C
34661-36857	23	36054	G	G	G	A	A	G	G	G	A	G
34661-36857	24	36207	C	C	C	C	C	C	C	C	C	C
34661-36857	25	36330	G	G	G	G	G	G	A	G	G	G
34661-36857	26	36617	C	C	T	C	C	C	C	C	C	C

(a) Region examined represents the nucleotide positions defining the start and stop positions within SEQ ID NO:1 of the regions sequenced;

(b) PS = polymorphic site;

(c) Position of PS within SEQ ID NO:1;

(d) Alleles for POR isogenes are presented 5′ to 3′ in each column;

45

Region Examined(a)	PS No.(b)	PS Position(c)	Isogene Number(d) (Part 2)									
			11	12	13	14	15	16	17	18	19	20
3878-4429	-	-	-	-	-	-	-	-	-	-	-	-
22306-22748	1	22635	G	G	G	G	G	G	G	G	G	G
29363-29820	2	29577	C	C	C	C	C	C	C	C	C	C
29363-29820	3	29754	C	C	C	C	C	C	C	C	T	T
30215-30747	4	30427	C	C	C	C	C	C	C	T	C	C
30942-31856	5	31531	A	A	A	G	G	G	G	G	A	A
30942-31856	6	31644	C	C	C	C	C	C	C	C	C	C
30942-31856	7	31648	C	C	T	C	C	C	C	C	C	C
32039-32600	8	32446	A	G	A	A	A	A	A	A	A	A
32039-32600	9	32478	C	C	C	C	C	C	C	C	C	T
33448-34167	10	33538	G	G	G	C	C	C	C	C	G	G
33448-34167	11	33551	A	A	A	G	G	G	G	G	A	A
33448-34167	12	33571	T	T	T	C	C	C	C	C	C	C
34661-36857	13	34766	T	C	C	C	C	C	C	C	T	T
34661-36857	14	34796	C	C	C	C	C	C	T	C	C	C
34661-36857	15	34797	C	C	C	T	T	T	T	T	C	C
34661-36857	16	34850	G	G	G	C	C	G	C	C	G	G
34661-36857	17	35325	G	G	G	G	G	G	G	G	G	G
34661-36857	18	35631	C	C	C	T	T	T	T	T	C	C
34661-36857	19	35632	T	T	T	G	G	G	G	G	T	T
34661-36857	20	35721	C	C	C	C	T	T	T	C	C	C
34661-36857	21	35774	T	T	T	C	C	C	C	C	C	C
34661-36857	22	35992	C	C	C	C	C	C	C	C	C	C
34661-36857	23	36054	G	G	G	G	G	G	G	G	A	A
34661-36857	24	36207	C	C	C	C	C	C	C	C	C	C
34661-36857	25	36330	G	G	G	G	G	G	G	G	G	G
34661-36857	26	36617	C	C	C	C	C	C	C	C	C	C

75.

(a) Region examined represents the nucleotide positions defining the start and stop positions within SEQ ID NO:1 of the regions sequenced;

(b) PS = polymorphic site;

(c) Position of PS within SEQ ID NO:1;

(d) Alleles for POR isogenes are presented 5′ to 3′ in each column;

80

	Region Examined(a)	PS No.(b)	PS Position(c)	Isogene Number(d) (Part 3) 21
80	3878-4429	-	-	-
	22306-22748	1	22635	G
	29363-29820	2	29577	T
85	29363-29820	3	29754	C
	30215-30747	4	30427	C
	30942-31856	5	31531	A
	30942-31856	6	31644	C
	30942-31856	7	31648	C
90	32039-32600	8	32446	A
	32039-32600	9	32478	C
	33448-34167	10	33538	G
	33448-34167	11	33551	A
	33448-34167	12	33571	C
95	34661-36857	13	34766	T
	34661-36857	14	34796	C
	34661-36857	15	34797	C
	34661-36857	16	34850	G
	34661-36857	17	35325	G
100	34661-36857	18	35631	C
	34661-36857	19	35632	T
	34661-36857	20	35721	C
	34661-36857	21	35774	C
	34661-36857	22	35992	C
105	34661-36857	23	36054	A
	34661-36857	24	36207	C
	34661-36857	25	36330	G
	34661-36857	26	36617	C

110   (a) Region examined represents the nucleotide positions defining the start and stop positions
within SEQ ID NO:1 of the regions sequenced;
(b) PS = polymorphic site;
(c) Position of PS within SEQ ID NO:1;
(d) Alleles for POR isogenes are presented 5´ to 3´ in each column.

115

POLYMORPHISMS IN THE POR GENE

```
CAGTGGCACA ATGTCAGCTC ACTGCAAGCT CTGCCTTCCG GGTTCACACC
ATTCTCCCAC CTCACCTCCC TCCTCCTGTT TCTGGGACTA CAGGTGCCTG       100
CCACCACGCC TGGCTAATTT CTTGTATTTT TTTTTTTTAA GTAGAGACGG
GGTTTCACTG TGTTAGCCAG GATGGTCTCG ATCTCCTGAC CTCGTGATCC       200
ACCCGCCTTG GCCTCCCAAA GTGCTGGGAT TACAGGCATG GGCCACCGTG
CCCGGACGAC CTGACTTACT TTTTAAAGCA TCCCCCTGTC TGCTGTGTGG       300
GGAATGGCCC GTCCATCAGA ATTCTCTTGC GGTGATTACG TAGGTGCAGA
TGTACCGCTA ACCTCACTCC CAGCGTCGTC GATTTATTAA TAGAAACTGC       400
CATGAGTTAT GCAATCAGCT TGAGTCTGCG TTGAGAAGGA CTTCACATCA
TGGGCATTAG ATCTCTTCTG GAAAATGAGG GGACCGGTCA CTTGGGGAGT       500
TAGAAGGGAG GAGGGTGAGG GACAGGGAGC GCGAGTTTCT TGGCTGCTGC
CTTGTGATGG ATAACAGCTA GACAGCTAGG GGAGCTCCTG TGTCACCATG       600
GCACTCCTGA CTGGTTCCAG TGGGGCCCTG ACTGTCTGCT GCCTCATGAC
TCTGTGGGAC AAGTCCTTCC AGAATCTGTC AGAGCAGATC CTAAATGCCT       700
CTCTGTCTGG TTCTATGAAA ATGTGACAGG TGGCTCATAA GCTTGACAGA
GATCCTGATA TGACTAGCAT TAAGTAAATA TTAAATGTGT TTATAAGATT       800
AAATATAAAT AGTTTTCTCT GGGGACAACA GAATTCACTT TGTCCTTTAA
TGTCATATCA AAAAAGCTGG CCAGGTGCAG TGGCTCACGC CTGTCATCCC       900
AGCACTTTGG GAGGCCGAGG CAGGATCACG AGGTCAGGAG TTTGAGACCA
GCCTGACCAA CATGGTGAAA CTCCATCTCT ACTAAAAATA CAGAAAATTA      1000
GCCGGGTGTG GTGGCACATG CCTGTAATCC CAGCTGCTCA GGAGACTGAG
GCAGGAGAAT TGCTTGAATC CGGGAGGCAG AGGTTGCAGT GAGCCGAGAT      1100
CGCACCACTG CACTCCAGCC TGGGTGACAG AGCGAGGCTC TGTCCCCCGG
CCCCCCCCCC CCCCGAAAAA AAAAAGTCA TATCAAAAAA GCCAAGTTGA      1200
GGCCAGGCAC AGTGGCTCAC GCCTGTAATC CCAGCACTTT GGGAAGCCGA
AGCAGGCAGA TCACCTGAGG TCAGGAGTTC GAGACCAGCC TGGCTGACAT      1300
GGTGAAACCC CGTCTCTACT AAAAATACAA AATTAGCCAG GTGTCGTGGC
GGGCACCCGT AGTCCCAGGT ACCCTGGAGG CTGAGGCAAG AGAATCGCTT      1400
GAACCCAGGA GGCGGAGGTT GCAGTGAGTC GAGATGGTGC CATTGCACTC
TAGCCAGGGT GACAGAGAGA GACTCTATCT CAAAAAAATA AAATAAAAAA      1500
GAAAGAAAAA AGCTAAATTG AGATGAACTC AAGCCATTGG AAAAGCAGCT
CAGATTAGTG AGAAAGCAAA GACCCAAGCT TGCGGGCCTG TTCTGAATGG      1600
CCTGTTCCCA GCAGGTTGTG TAGCAGTAGA CCGATGCCCA CAACCCTCGG
ATGCCCTGGC TTCCCAGACC AGCCCCATGG TTTCCTGGTG ACAGTGACAG      1700
CCATCAGCAG TCAGCTAGAA CATAGTCTCT CATTGCCTAA TTGCTGACAC
ATGACTAAGG CTTTCCTGAT TGAGTCTGGG TGCTCTGGTT GTGGTGAAAA      1800
ACCCAGTCGG TCATATCCAT GATTGCAAAA TGGAATAGTG TGAATTTCAG
GGCTGAGCCC AGCCTCCTTG GTAAGCAGCA GAAGCTGCTG TTTCAGGGAC      1900
TCTCACTTGG ACCTTTGTTG ACTCATGTCA CGTCTGCCTA CCAGGTGTCG
TTGGTGGAAG GACCAGTTGA GCATCTTATG AGGGAACTAG GCTGGTGACA      2000
TGGGGGTTCC CCCACCTCCC TCCTGGACGG GGCGGCTGGC CGGGCGGGGG
GCTGACCCCC CCACCTCCCT CCCGGACGGG GCGGCTGGCG GGGCAGAGGG      2100
GCTCCTCACT TCCCAGTAGG GGCGGCCGGG CAGAGGTGCC CCTCACCTCC
CGGATGGGGT GGCTGGCCGG GCGGGGGGCT GACCCCCCCA CCTCCCTCCC      2200
GGACGGGGCG GCTGGCCGGG CGGGGGGCTG ACCCCCCCAC CTCCCTCCCG
GACGGGGCGG CTGGCCGGGC GGGGGGCTGA CCCCCCCACC TCCCTCCCGG      2300
ACGGGGCGGC TGGCCGGGCG GGGGGCTGAC CCCCCCACCT CCCTCCCGGA
CGGGGCGGCT GGCCGGGCGG GGGGCTGACC CCCCCACCTC CCTCCCGGAT      2400
GGGGCGGCTG GCCGGGCAAA GGGGCTCCTC ACTTCCCAGT AGGGGCGGCT
```

FIGURE 1A

```
GGGCAGAGGC GCCCCTCACC TCCCGGACGG GGCGGCTGGC CGGGCGGGGG        2500
GCTGACCCCC CCACCTCCCT CCCGGACGGA GCGGCTGGCC GGGCGGGGGC
TGACCCCCAC CTCCCTCCCG GACGGGGTGG CTGCTGGGCG GAGACGCTCC        2600
TCACTTCCCA GATGGGGTGG CTGCTGGACG GAGGGGCTCC TCACTTCTCA
GACGGGGCGG TTGCCAGGCA GAGGGTCTCC TCACTTCTCA GACGGGGCGG        2700
CCGGGCAGAG ACGCTCCTCA CCTCCCAGAT GGGGTCGCGG CCGGGCAGAG
GCGCTCCTCA CATCCCAGAC GGGGCGGCGG GGCAGAGGCG CTCCCCACAT        2800
CTCAGACGAT GGGCGGCCAG GCAGAGACGC TCCTCACTTC CTAGATGGGA
TGGCGGTGGA GAAGAGGCGC TCCTCACTTC CTAGATGGGA TGGCGGCCGG        2900
GCAGAGACGC TCCTCACTTT CCAGACTGGG CAGCCAGGCA GAGGGGCTCC
TAACATCCCA GACGATGGGC AGCCAGGCAG AGACGCTCCT CACTTCCCAG        3000
ACGGGGTGGC GGCCGGGCAG AGGCTGCAAT CTCGGCTCTT TGGGAGGCCA
AGGCAGGCGG CTGGGAGGTG GTTGTAGCGA GCCGAGATCA AGCCACTGCA        3100
CTCCAGCCTG GGCACCATTG AGCACTGAGG GAACGAGACT CCGTCTGCAA
TCCCGGCACC TCGGGAGGCC GAGGCTGGAG GATCACTCGC GGTTAGGAGC        3200
TGGAGACCAG CCCGGCCAAC ACAGCGAAAC CCCGTTTCCA CCAAAAAAAT
ACGAGAACCA GTCAGGCGTG GCGGCGCGCG CCTGCAATCG CAGGCACTCG        3300
GCAGGCTGAG GCAGGAGAAT CAGGCAGGGA GGTTGCAGTG AGCCGAGATG
GCAGCAGTAC AGTCCAGCTT CGGCTGGGCA TCAGAGGGAG ACCGTGGAAA        3400
GAGGGAGGGG GAGGGGGAGG GGGAGGGACC AGACTTTGTC TCTTAAAAAA
AAAAAAAAAG ACTCCCAGGC ATAGGAAAGC AGGCATTCAC CATAAATCTC        3500
ATTGTTAACA GGAACTACCT TCATAGCCCA AGGTTTTAGG AATACAGAGA
CACTCTTGCC AGTCAGGATA GTCCAAGGCT CGTGAATATC AGGAGCCACT        3600
CAAGGGCCAG GCCTGGAGAC TTGGAACATG CAGGGTTTAG GGCAACCCCC
GCCAGCCACA CAGCTGGGGA AAGGGTTTGC TGCGAAATGG TCCACTCTTG        3700
CGTGCACTTC CTCCCGCCTG CCCCCCGTGC GGTGTCCTTG TGGGATTAGA
GGATGTTCCA GCACACTGAG AGTCAGTGGC TGTCATTTTC TCTGTCTTCC        3800
AGGGACCCAG TCCCAGCTGC TGCCTTAGCT TGCACCCCAG GACCACCCTG
GGCTGGTGCT GACACCACTG GGGGCTGTCT GTTATGTCAG CCCCAGTCCA        3900
AGACTGTGGC TGTCTTGGCA GGAACCTGGC TGAGTGAGCC CCTTCTCCTA
CCCCGTGCAG TGACCATTTC CTGCAGCCCC AGGGGCAAGG CCCAGCATTT        4000
AGGTGGGCAC CCCAGCCAGC CTCAGGGGCA CCCTGCTACC CTCTGCTGAC
ATCTGCTGTT TCTGTCTCCT AACAGTTTCA TGATCAACAT GGGAGACTCC        4100
     [exon 2: 4092..
CACGTGGACA CCAGCTCCAC CGTGTCCGAG GCGGTGGCCG AAGAAGTATC
TCTTTTCAGC ATGACGGACA TGATTCTGTT TTCGCTCATC GTGGGTCTCC        4200
TAACCTACTG GTTCCTCTTC AGAAAGAAAA AAGAAGAAGT CCCCGAGTTC
ACCAAAATTC AGACATTGTA AGTGCCGCCT CTCAGCCTCC TCTCTCTGTC        4300
          ..4267]
CCTCTTCTGT CACCCACTCCA AGCAGTGTCC TGCATGCGGG CCTCAGGCTG
AAAGAAGCAA GGCTCCTTGT AGCATTTTGG GGTGACTCTT GGGTTTTGCC        4400
TTCTTGCTAG ATTGTGCTTG AGCTCTCCCT TAATCATCTC TTATGTGGGG
TTTATAAGGC CCTTTTCTAT CTGCTGTTTA TATGACCTTC ACAGGAAGCA        4500
CTGCGAAAAT GACACCAGTT TACAGATTAA GAACCGGGGG CTCAAGAATG
TTAACGGGCC TGCCAGGGTC ATTAGCCAGC AATAGCAGCG CTAGAGCTGG        4600
AACCCTTTTT TCCATTTATT TATTTATTTA TTTATTTGGA GACATGGTCT
CACTGTATTG CCCAGGCTGG AGTGCAATGG CGGTGGCACG ATCAAGGTTC        4700
ACTGTAATCT CCACCTCCCA GGCTCAGGCA GTCTTCCCAC CTCAGCCTCC
CGAGTAGCTG GGACTAAAGG TGTGCACCAC TATGCCCAGC TAATTTTTTT        4800
TTTTTTTTTT TTGAAGCAGA ATCTTGTTCT TTCAACCAGT CTGGAGTACA
GCGGTGCGAT CTCGGCTCAC TGCAACCTCT GCCTCCCAGG CTCAAGCAAT        4900
TCTTCTGCCT CAGCCTCCTG AGTAGCTTGG ATTACAGGTG CCCGCCACCA
```

FIGURE 1B

```
CGCCTGGCTA TTTTTTGTAT TTTTAGTAGA GACAGGGTCT CACCATGTTG      5000
GCCAGGCTGG TCTCAAACTC CTGACCTCAA GTGATCCGCC CACCTCGGCC
TCCCAAAGTG CTGGGATTAC AGGTGTGAGC CGTTGCGCCC AGTCAATTTT      5100
TGTATTTTTT TATAGAGACA GGGATTTGCC ATGTCCCGGG CTGGTCTCGA
ACACCTGGAC TCAAGCAATC CGCACACCTT GGTCTCCCAG AGTGTCGGGA      5200
TTACAGGCGT GAGCCACTGT ACCCAGCCTA GAATTGGCAT TTTAAAGTCT
TTTTAAGGCT GTAGCCACCT GGAAGGCTAT GTGCTAATGA GCACAAAATC      5300
TTTGGACAGG CCTTATTTTG TTGCTCTTTT AAACTTTTCA TGAAGCTTCC
TTTTAAGAAT AAGTGAAAAA GATGGATTAG TTAAAGAAA AACGTTAAGG       5400
AAATAATAGT ATAGATGGTA TGTAGACAGA AAAAAATGAT ACCGAAGGTC
TGCCTAAGTG ACTGCATCCC TGTCTCTCCA AGATTGATTT CCATAAGCCA      5500
CACGGAAAAG TCATTCTCCA GATTTTAGGG TCCTGCGGCC AGAATTACTT
CCTGCGGTTA CGCTTGGCTG TGTTAACAC CAGTCTCAGA CTCCAGGGGC       5600
CCGCAGCCAG GTTGAGTGG GGACCTGCCA GGTCAGGTCT GAGCCGAGGG
AACTCCAGGG GCACGGGCCT GCTGCCTTCA GGGACTCCCT GGATTTGGCA      5700
AGGGGCAGGT CTCAGGGGCA GGATTGGGAG CCATTCTCAG ACCAGAACCC
CTTTCATTCT GTCTGTCTTC AGGGGACCCA GTCCCAGCTG CAGCCTTAGC      5800
TTGCACCCCA GGGCCACCCT GGGCTGGTGC TGACACCAGT GGGGGCCATC
TGTCATGTCA GTTCCAGTCC AAGGCTGTGG CTGTCTTGGC AGGAACAAGC      5900
CACACAGAAA AGTCATTCTC CAAGGCTGGG CATGGTGAAG CATGCCGCTC
ACATCAAATC CCGATTCCCT ATTCCAATCT GGGTATCAGC CTTCCCCGCC      6000
AGAACACTTA CAGCCTTCAC ACCGTAAAGC AGATATTCAC ACAACACAGT
TCTTATACCA TGGACAGGGA ATTAGTTTTA CATTTTTAAA AACTGGGTTC      6100
AGGGCCAGGC ACAGCGGCTC ACGTCTGTAA TTCCACCACT TTGGGAGGCC
GAGGCGGGAG AAATCACTTG AGGTCAGGAG TTCGAGACCA GCCTGGCCAA      6200
CATGGCGAAA CCCCGTCTCT ACCGAAAATA CAAAAACTAG CTGGGCGTGG
TGGTGCATGC CTGTAATTTC AGCTACTAGG GAAGCTGAGG CAGAAGAATC      6300
GCTTGAACCC AGGAGGCAGA GGTTGCAGTG AGCCAAGAAT GCGCCACTGC
ATTCCAGCCT GGGAGACAGA GTCCGTCTCA AAAACAAAAC AAAACAAAAC      6100
AAAACTGGGT TCAATCTTAT CTCCATCATT GACTCACTGA GCAAACATTT
CCTGAGGCCC TACCTACAGT GTCCCAGACA CTATGCTAAG GTGTGTGGCT      6500
GCAGCAGTGA AGAAAATAGA TGAATGCTGT CTTCATCAAG CTTACGCTCT
AGTGGTGAGT GAACCCTGGG CGAGACCTTA ACCTCATTCT GCCTCAGTTT      6600
CCTCACCTGC TAAATGGAGG CAGTATTCCC AATCTTGTGG GGCTTTTGTC
TTAAATGAGG AAGAGTCTAT AAACTCACCT AGCACTGCCC CTGAGGGATA      6700
GGAGCTGTGT AGTGTTTGCG TTGATCGTAT TTCGTGCCAG GCACTATGAC
GGGCACGATC AGGACCACAG AGCCAGCATT TATGTTAAGC TGGAAAATGA      6800
GGGTGTTGCA ACATAGTTTT TTTTTTTTG AGACAGACTT TTGCTCTTGT
CACTCAGGCT GGAGTGCAGT GGTGCAATCT CGGCTCACTG CAACCTCCGC      6900
CTCCCGGGTT CAGGCGATTC TCGTGCCTCA GTCTCCCGAG TGCCTGGGAC
TACAGGCACA TGCTACCACG CCTGGCTAAT TTTTGTATTT TTAGTAGAGA      7000
CGGGGTTTCA CCATGTTGGC CAGGCTGGTC TCAAACTCCC GACCTCAGGT
GAGCTGCCTG CCTCTGCCTC CCATAGTGCT GGGATTATAG GCGTGAGCCA      7100
CCACACCCGG CCACAGCGTA CTTTTTAGAA ATGCTTTGTG AACTGCTTTA
ACCTACATTG TCTAATCAGG GCCTCCTTGC AACAGCGCCG TGCGTTAGGT      7200
AGCGTGGGTT GCCTCATTTC TGCTCTCAGA GAGGTTATGG GATGTGCCCA
GGGCACAGGC TGGAGAGACA GGCTCCGCCA GCTCTAAGAG AGAAGGAATT      7300
GGTGTGGGGA TCAGACCAAG TCCTGCAGGA GGGGTGTTGG GCAGGTTCTT
AAAGGTAGAG CCACACCCTG AAGCTGAAAG GACTTAGAG CAGAGGGCCC       7400
GATATAGGAA GGGCAGATCA GCTGGAATCC AGCCCATCCT GTCCAGGGCT
GGTACACGGG GCATTGCTGG TGTGGAATAG TACTTTTGAT AAAAAAGCGT      7500
AGTAGTACTC AAGATAAAGA GACCAGGGTC CAGGACCCAG CTCCTGAACT
```

FIGURE 1C

```
GGCCCTTGTT  CCCTCATCTG  TAAAATGGAA  GTTCCGCCCC  ACCGCCGTGA     7600
GTGAGGACGT  GGGTTGGCCC  TGGGAGCCTG  CGGAGCTTGC  CACCAAGGGA
GGCAGCATTC  CTCTCATAGC  TCATCTGTTC  TCGCTGTCAC  CTCCCTGGCC     7700
TTACAAGAGG  AAGTTGACAT  AAAACCTGTA  GACTGTGGGA  GCTGTAGGGT
CCCAAAGCAC  TTGTATTTTT  TTGTTGATTC  TCACTGTTCC  TCTTTGCCAG     7800
GATGCTGAAA  TTCTGTTCCC  TCCACCTAAT  CATGATCTTC  TAACCTTATG
TGTGTGGAGA  TATGCAGTTT  CCAGAGTCTT  GTCTCTTCCC  TCATTAAATC     7900
ACACTGGATG  CCTGCTGCAG  CCAGGGCTTT  TACCTGTACG  CTGTACAAGA
CAAATATTTC  TAGTACCATG  GCTACATTCT  GGACCCTGAA  GTTGGGTGCT     8000
CAAGGTTGAG  TACTGACTCT  CCCACATTAG  CCATCTGACA  TCAGCCAGGT
GACTTGATCT  TTCCAAGTCT  GTGTCCCCAA  CTGTAAGCTG  GAGGTAGTAG     8100
TAGTACTTAC  CTCATAAGGC  TGTGAAAACT  GAGTTAATCC  ATCATAGAGC
TCTTTGCTCA  GTTCCTGGTG  CATAGTAGGC  TCTCAGGTGT  TATGGATTGT     8200
TACTATTTAT  TTATTTATTT  TGAGGCAGGG  TCTCACTCTG  TCGTTCAGGC
TGGAATACGC  TGGCACGATC  TCGGCTCACT  GCAACCTCCG  CCTCCCAGGC     8300
TCAAGAGATC  CTCCCACTTT  GGCCTCCCAA  GTAGCTGGGA  CTACAGGCGT
GCACCACCAC  CCTTGGCTAA  TTTTTATATT  TTTTGTACAA  ATGGACATCT     8400
AACAATGTTG  CCCAGGCTGG  TCTGAAACTC  CTGGGCTCAA  GCAGTCTGCC
TGCCTCGGCC  TCCCAAAGTG  CTGGGATTAC  AGGCGTGAGC  CACCATGCCT     8500
GGAGTGTGTT  ATTATTTTTA  TCCTTTTTAT  TAATACCATC  TCCAGTCTTA
CAAGAGCCCT  TCCACGAAGA  CATGAGCAGC  CCATTTTGCG  GATGAAGAAA     8600
CCGAGGTGCC  AGGGATTGCA  GAGTTTGCCC  AAGGCCACCT  AACTTCTAAA
TAGCAGCAGA  GTTGGCTTTT  GAATGAAGTC  TGCAAGCTCC  ACAAGCCATG     8700
TTTTCTCTAC  TTTGCCATGC  TGCTCCTAAT  TTAGTGTTAA  TAGGCCTCCT
ACGATAGTGT  TAGTCTCCTA  AGATAGTGTT  AATAGGCCTG  GCATGGTGGC     8800
TCACGCCTGT  AATCCCAGTG  CTTAGGGAGG  CTGAGGCACA  AAGATTGCTT
GAGCCAGGAG  TTCAAGACCA  GCCTGAGCAA  CATAGGGAGA  CCCCCACCCC     8900
CCCGCCATCT  CTACAATAAA  CAAAAAAGAG  TGGTGTTAAT  TTGTTAATTT
ACCGACAAGG  AGAGTATTCC  TTATCTGAAG  TGCTTGCCAG  AAGTGTTTCA     9000
GATCTTGGAC  TTTGGAATGT  TTGCATTATA  CTCACCAGAT  GAGCATTTCT
TATCTGAAAA  TCTGAAATGC  TCCAATGATT  CTTCCCTTTG  AGGATCATGT     9100
TGACGCTCAA  AAAGTTTCAG  ATTTCAGATT  TTAGATTAGG  GACATTCAGC
CTGTAGATGT  TATTTTACAG  AGCAGGAAGT  TGGGTGTAGT  GGTTCACACT     9200
TGTGATCCCA  GCTCTTTGGG  AGGCCAAGGC  AGGAGGATTG  CTTAAGCCCA
GGAGTTGGAG  ACCAGCCTGG  GCAACGTAGC  AAGGCCCTGT  CTCTATAAAA     9300
AATAGGAGAA  AACTGAGGCT  CAGAAAGGTC  ACCTAGCTGA  TACAACCAGG
CCTCAAGCCT  AGATCTCCTA  TTCTAAATCC  CACATCAGTG  TCCCTGGCTA     9400
CCGGTATTTC  AGTGAGCACA  GGTGATTTCT  GTTATAGCCA  CAGAGCAGTA
GGGAGTAATA  GGGCTTAGTG  TGAGGAAGGT  ATGTGGCCCA  GGGCTATATC     9500
AGGAACTCTC  ACCATGATTT  CTGGTGGCTT  CTAGAATTGG  CATTTATTTA
TTTTGTTTTA  GAGACACGGT  CTCGCCCTGT  CACCCAGGCT  GGAGTGCTAT     9600
GGCGCGATAT  CAGTTCACTG  CAACCTCTGC  CTCTGGGGCT  CAAGCGATTC
TCTTGCTTCA  GCCTCCCAAG  TTGCTGGGAT  TACAGGCACC  CACCATGATG     9700
CCCGGCTAAT  TTTTGTATTT  TTAGTAGAGA  TGGAGTTTCA  CCACGTTGGC
CAGGCTGGTC  TTAAACTCCT  GACCCCAAGT  GATCCATCCG  CCTGGGCGTC     9800
CCAAAATACT  GGGATTACAG  GTGTGAGCCA  CTGTACGGGG  CCAAACACAT
TTTTAAAACA  AAAAACCCAG  TGGGATTTTT  GTTGGAATTG  GATTGAATCT     9900
GTATATTAAT  ATGGGAAGAG  TTGACATCCT  ACAATATTGA  GCTTTCTATT
CCATGAACAT  TGTTTTATCC  CTCCCGGGGA  GGTCTTTTGG  ACTGTTTGTA    10000
ATAGCAATTA  ATGACTTGTT  AGAGAGGAGT  GGTAATGAGC  AACAAGTTCT
TTTTCATAAT  AATAGTAGCA  GTAATGATAA  TAAATGGCTC  TTCCTGCCAG    10100
GTGTGGTGGC  TCACGCCTGG  AATCCCACAC  TTCAGGAGGC  CGAGGCAGGT
```

FIGURE 1D

```
GGATTGCTTT GAGCCCAGGA GTTCAAGATC AGCCTGGCAA CATGACGAAA        10200
CCCTGTCTCT ACGGAAAAAA TACAAAAATT AGCCAGGCAC GGTAGCATAC
GCTTGTAGTC CCAGCTACTC AGGAGGCTGA GGTGAGAGGA TCACCTGAGC         10300
CTGGGATGCC AAGGCTGCCG AGCGCCGAGC TCACACCGCT GCACTCCAGC
CTGGGCGACA GAGCAAGACC CCCTCTCAAA ATAATAATAA TAATAATAAT         10400
AATAATATAT AATAATGACT CTTTCACGTC CACTCCCACT CACTCAGGTT
GCACCTCTGA ACAGAGCATT TCGCACTGGG GCAGCTCATG CCCTAAGGGA         10500
GCCCCAGGGA ATGCTCAGCC TCTTGAGTAG CTGGGGTGAC GGGCACATGC
CACCATGCCT GGCTTATTTA TTTTTTTTTA AAGACATGGT CTCGCTCTGT         10600
TGCCCAGGCT GGGTCTTAAC CTCCTGGCCT CAAGCAGTCT CCCTGCCTCA
GCCTTGCAAA GTTCTGAGAT CACTCACTGT GTAGGATCCA AAGTCCCACA         10700
GGATCCAGGG CTCCTGTGGG CTCTCTCTCC TGTGGGCCTG GCAGCTTTGC
TCACCCTTAT GGAACAACAC CGCATGGCGT GCCCTCTTGG TGATGGAATT         10800
TGTATTTTTG CCTCCCATGG TGCAGAGAGC GTCCCATTTC CATCTGGGTC
CCTACCTTAG TGCGGGGCCG CCTGTCGGAG GGAAGCTTCT CAGAGAATGG         10900
CCGTTGAATT AACCAAGGCT AAATCTGTAT GTGTGGCTGC CTCTAGGGAA
ACCTGTGGCC TCCAGGCTGG GTTTGGCTTA CACAGTATTT TTTTAAAAAA         11000
TATTTTAATT AGAACATTAA AAAGTGTGGC AGTATCAGGC CCAGCGTGGT
GGCTGACACC TGTAATCCCA GCACTCTGGG AGGCCAAAGC AGGTGGATCA         11100
CATGAGGCCA GGAGTTCAAG ACCAGCCTGG CCAACATGGC AAAACCCTGT
CTCTACAAAA AGTACAAAAA TTAGCTGGGC ATGGTGGTGC GTGCCTGTAA         11200
TCCCAGCTAC TCGGGAGGCT GAGGCAGGAG AATTGCTTGA ATCCAGGAGG
TGGAAGTTGC AGTGAGCTGA GATCACGCTA CTGCATTCCA GCCTGGGCGA         11300
TGAAGTGAGA CCAAAAAAAA AAAGGCAGTA TCAAATAAAA ACTTAGACTT
ACAGCTTCTT TAAAAAAAAA AAAAAAAATA GAGGATCTGG TCATACTGGA         11400
CCCACAGTGA TGCCCACCAG CTGTGGAGCA GCTGAGCAGC GGCCCCATTC
TGCACGAGGC CTGTGCGCTG CATTCACTCT GACCCTCCAG CCCGCTCAGG         11500
CTGCTGTGCT CATGTTGCCT CTACTGGCCC TGAGAAGCTC TGCAGTGCCC
CCTGCACCTG TGCTTTGCTC CTGGAATGCT AGAGTGCTTT AAAGAAAAAT         11600
AAAGCTTCTT ATGGAACATT TCAAACACCA GAGTGTAGGG CACAGGAGAG
TGGACCCCAT ACGCCTCCAG CCCCCAGTTT TCATAACTAC TCACTCCTCG         11700
CCGGTCTTAT TTCTTTCCGC TGCCCCCCAC CGCCGCACCA GTCCTGGATT
ATTTTAATGC AATCCCCAGA TGTCATATTT CATCAGGAAA TGTTTCAGTC         11800
CTCTAAAATT CAAGGACTCG CCAGGCATGG CGGCTCATGC CTGTAATCCC
AGCACTTCGG GAGGCTGAGG CAGGAGAATC ACTTGAAGCC AAGTGTTCGA         11900
GATCAGCCTG GGTAGCATAG CAAGACCCCG TCTCTACAAA AAACACAAAC
ATTAGTCAAG TGTGGTGGCA CATGCCTGTG GTCCCAGTTA CTCAGAAGGC         12000
TGAGGTGGGA GGATCACTTG AGCCCAGGAG GTTGAGGCTG CAGTGAGCCA
TGGTTGTGCC ACTGCACTCC AACCTGGGTG ACAGAACGAG ACCCTGTCTC         12100
AAAAAAAAAA AAAAGTAAAC TTCAAGTATT CTCCTTTCTG AACATAACAC
AACGCCATCT TCACACCTTG GAAGTCAACA GTAATTCTTC AGCATCGTTA         12200
GATACCCAGA CAACATTTAT TCATAGTTCC CTGATTGACT GATGTTTGAT
TTTGTTTAAC AGTTTTTGTT TGTTTAATTC AGGATCTAAA CAAGACCCAT         12300
ACACTGCAGT TGATTAATAA ATCTTTCGTT TGGGAGAGGT CAGGTATATT
CAGGTATAAT TTGCATACAG TAAAATGCCC CCCTCCTCCC TTTTTTTTGA         12400
GACAGGGTCT TGCTCTGTCA CCCAGGCTGG AGTGCAGGGG TGTGATCAGG
GCTCACTACA GCCTCTACCT CCCAGGCTTA AGGGATCGT CCCATCTCAG          12500
CCCCCTGAGT AGCTGGGGCT ACGGGCATGT GCCACCATAC TCGGCTAATT
TTTATTTTTA ATTTTAGTAG AGACAAAGTC TCACTATGTT GCCCAGGCTG         12600
GTCTTGTTCG AACGTCTGAG TTCAAGTGAT CCTCTTGCCT CAGCCTCCCA
AAGTGCTAGG ATTAGGATTA CAGGCATGAG CCACTGCACT GGGCCAACAT         12700
TTTCCTCTTT AAGTGCACCA GTTCATGAGT CTTGACAGAT GTATGTGATT
```

FIGURE 1E

```
CTGTAATCAG TATCATAATC GAAATCTAGA ACATTTCATC ACTCCGAAAC    12800
AGTCCTGTCC TACCTCTTTG CAATCATCCT CCTCCATCGG GTAATCAGCC
TCTGGCAACT GCCAATCTTA TTTCTGTCCC TGTAGTTTAA CATTTTCCAA    12900
AGCGTCACAT AAATGGAATG ATAGCACCGT GGGCTGCTGT GCCAGCGGTT
TTTCCTTAGC GACACGTCAG CTCCCTTTCC ATTCCTAAGC TTCCCTTCCA    13000
TCCCGTTTCT TCCTTGAAAC ATCATTTGTA TCACAGTTTC CCACAGTCTG
GATTTTCCTG ATTACGTCCC CCTGGAGATT GCTTTTGAAA AGCCACAAAA    13100
GATCTCATTT TCCTGAGTGA CCTTCCTTTT GGGGCCTGTT CTTGACAGAA
AGCCCGGGAG GCCAGGGGTC AGGGCTGCCA GCATCACGTG TCTGTTTTTT    13200
ATCTCCTGAG TGCCAACAGA TGGGAAGTGG ATCTGGGGGT CTTGCCACAT
CCAGTTCAGG GTCACTTTCT CTGTAGTGAG GTCCTTGGCT CCATCGAGCA    13300
TGAAAAGAGA AATACAAATG AGACTGACAA AGAGGAAAGC ATTAGGGGAA
GCCTGTGGCC ACGAGAGTAG ACGGTGCAGG GGGCCCACTG CCTCCCGCTG    13400
CTGTGGGAAC GGCGAGGACA TGGGCCAGGG TGTTTCTGCC ACTTCTGCAA
GCACAGCTCT GGGTTTTCTG CTGCCATCTG ACCTCCAAAG GGACCTCGCC    13500
TCCTCTCCTT CCTTTCCTGG TCCCTGATGA GATGGGGGAG GTTTCTGTCA
CTCTGACAAG TGCTTGAAAC TCCAGACATG GGCTGTCGGG AACTGTCGGG    13600
TACTGCCGTG CAGAGCGCAT GTTGTCTGAC CTCCTGACTT CACTGGGCCC
TGTCGGGTGC AGGAGAGGGA GCTGTGGATG GCACAGTGGG ACATGTGCTC    13700
TACTGGCCTG GGGGTGGTGG GAATGCAGAG GAGGGGTATC CCCATTGGGG
GTGTGTCAGG GATGAGATGC CCCCACAGGA GCCTAGGAGG CCAAGTGTGG    13800
AGGCTGGGAG GTGAGGACAC GCGGGGCTGT TCCAAGGGAG CCACATGCTG
GTAGAGCCCA GGCACAGGTC ACCAAGGGAG TCAGGGGTCG TGGGGGGTCA    13900
GAGGTCATAA AAGGAGTTGG GGTCACCAAG GAAGTCAGGT CATGGAGGGA
GCTGACACAG GAGAGCATCT TAGAGGGCCT TGGGACCTTC CAAGGGGTAT    14000
GGATTTTTTT CCCCAGCTTT ATTGAGGTAT AAATGACAAA AATTATATAT
ATTTATCAGA GCATCTAAGT TTTTTTCTTT TTCTTTCTTT CTTTTTTTTT    14100
TTTGAGATGG AGTCTTACTC TGTGGCCCAG GCTGGAGTTC AGTGGCACAA
TCTTGGCTCA CTGCAACCTC TGCCTCCTGG GTCCAAGTGA TTCTCCTGCC    14200
TCAGCTTCCT GAGTAGCTGG GACTACAGGT GCCTGCCACC ACGCCCAGCT
AATTTTTGTA TTTTGAGTAG AGATGGGGGTT TCACCATGTT GACCGGGCTG    14300
GTCTTGAACT CCTGACCTCA AGTGATCCAC CCACCTCAGC CTCCCAAAGT
GTTGGGATTA CAGGCGTGAG CCACCATGCC CAGCCTAAGC ATTTTTCTTT    14400
TTTTCTTTTT GAGACTAAGT CTCGCTCCTG TCCCCCAGGC TGAAGTGCAA
TGGCGCAATC TCGGCTCACT GCCACCTCCG CCTCCCGGCT TCAGGCGATT    14500
CTCCTGCTTC AGCCTCCCGA GTAGCTGGGA TTACAGGCGC CTGCCACCAC
ACCCGGCTAA TTTTTGTATT TTTAGTAGAG ACGGGGTTTC ACCATATTGG    14600
CCAGGCTGAT CTCGAACTGC TGACCTCAGG TGATCCGCCT GCCTCAGCCT
TCCAAAGTGT TGGGATTACA GGTGTGAGCC ACCGCACCAG GCCAGCATTT    14700
TTCTTTCTTT GTAAATTTTA TTTTTCCCAG CTTTAAAATG ACATAATTGA
CAAATAAAAA TTTTATACAT ATATTTAAAA TTATGTAAGT TCAAAGTATA    14800
CAGTGTGATG TTCTGATATA CGTATCTTAG CTGAGCATGG TGGCTTACAC
CTGTAATACC AGCTACTTGG GAGGCTGAGG CGAGAGGATC ACTTGAAGCC    14900
AGGAGGTCAA GGCCAGCTTG GGCAACATAG CAAGACATCG ACTCTACAAA
AACTTAAAAA TAAATTCAAA AATTAGCTGG GCATGGTGGT GTGGACCTGT    15000
CATCGCAGCA CTTTGGGAGG CTGAGGCAGG AGTTGGAGGC TGCAGTGAGC
TGTGGTCACG CCACTGCACT CCAGCCTGGG CAACGGAGCG AACCCCGTCT    15100
CTAGAAAAAC TAAAAAATCA ATTTAAAAAA ATGACATATG TGGATGACGC
TAACAGGGAG CTGGGGAAGC CTGTGTAGGA TGGCGGTGTC TGGAGAGCGG    15200
TGGCCCCATG AAGCGCCATT TGGAGCTGGA GTGGAGGGAG CTGACTTGGA
AGCAGGAAGA TGGGCCAGGA GGCTGCTGGC CTCAGCCTAG CAGAAATTGA    15300
CCAGGGCCGC TCGGCACAGA TGGAGAGCCG AGTGCAAATG CAGGTGTCAG
```

FIGURE 1F

```
TGGATCAGAG AAAGGGCTCC AAGCCAGTGC TTTTGAAAAC ATAATTCTGG      15400
GACCCTCATC ACCCACCAGA GGTGATGTTC TTATCTCCAT TTTACAAATG
GCAGTGCTGA TCCCAGGAGA GGGTGAGTGA CCTGGTCAAG GCTGACGTTA      15500
GCACATGAGA GCTGGGATTC CAGCCTGGGG CTTCCTGCAC CACGCTTTTG
CTCCCTCGGA CCCCTAGCTC TTGTCATGGC CACCCCAGCC CACCCCTCCT      15600
TTCCATCCCC TTTGATTGCA CACAATAGCA GAGCCACTTT GTGCCAGATC
AGATGATCTT ACTGTCTGAG GGTATGGCCT TTTCTTTCTG TCACCGATGG      15700
GCTGAATAAC CCTGAGTTTC TCTTCTCTGT CTCCGGAAAA TTAGCCCTAG
TCCTGGCCAC CCCCTGGCTC AGGGTCACTG CTGAGGTCGT CAGAAGAGCA      15800
TCAAGGCTTT TGCGTGTGTG GAACAGCACA GCAGGCTTTC TGAGCTTCAC
AGACTTGCTC ACGGAGTCCT GATGTGAACA CGTGCTGGAG GACATGGTCG      15900
GGGGGAATGA GGGAGCGGCA GCCCACCAGA GCCTTAGAGC ATCCTGAGGA
CACGGCCCCA GGATCCTTGG GAAGGGCTGT GAGATGTGCT GCCTGAACGT      16000
GCCCGGGGCG CACTGCAGAC AGGGTAAAGC CGGGGGAGGT GCACCTTGCT
TGGGGAGCAG ACAGGAACCC CGGGTGCTGG GCCACTGGGC CCTGGAGCTT      16100
TCACAGGCCT GGAGTTGAAT TACCGCTTCA CCCACCCTGA GCCTCACTGC
CTTGTCAAGG TGGATAGAAA CCCTTGCGCA CTGCACAGGG TGGTGAGGAT      16200
TAAACAAAAT AAAATAGAAA GCTTCTTGAA TGCTAGCCAG CATGCAGTAG
GCACTCAGTA AGAAGTCATT CCCTTCGCCT GTTTTGTCTC TGACCTGTGG      16300
CCACTCACAC CTGCTTTCAC CTCCCCCTGC CTGTAGCCAG CTGTCTCCAG
CAGCTGCCCT CCCCACGTGT TCATCAGAGT CCCTGTCCGC AGAGACAGCT      16400
GCTGGCTTCT GTCTGAGCTC CAGCAAACCA CTTTCTGTTT TCACCAAAGA
CTGGCGTTTC CAACCATTAG CCCCATGCCA CAAGGCTCTG AAGGTCAAGA      16500
AATCTGGGTC CGCCTAGCAG AGCCCACTCT CCGTCCTGCC AGAGCAGGCT
GCCCCCAGCT CTGCTGCCAT CGTGCCTCTG TGCCTCAGCC GCCACCCCTT      16600
CTCTCGCCGT GGGCACCCAT GACCCTGGAG CCTGGAGAAC TTGGCAGTAG
ACGCTAGGTT CTGAAAGGGA GCTGGTCCAA AGCAGCTTCC CCATGTGTTG      16700
CTAAGGGCAG GGGACAGTCT CAGGAGAAAT CGAGTCCGCT GCAGAAAATC
CAACCGGATT TGCAGTTTCC AGAAATGTTC CTCAGCGTCG CCTGTGACGG      16800
GAACAAAATG AACCGCACAT CTCGGTTCAT CTCGGTCCAA GTGACTCAGT
GCTTTTCCCG CCCTCCCTCA TCTTCCCCAC GCAGTGGACT ACAGCTGACT      16900
TCTGATCCTG AACCCAGAAA GACTCCTGCT CCCAATTTAG GCGTGGAGAG
CCTTAATTGG AACCTTTCAT GCTCCGTTCT GGAGCCACGT GGTCTGTGTG      17000
CTCCAGGCCT GGGACAGCTG CCACCTCTGT ACGGCTGGAG CCTTGGGCTT
TCTGGTTTTG TTTGTTTTTT CTTGAGACAG GGTCTCACTC TGTCAACCAG      17100
GTTGGAGTGC AGTGGCACAA TCTTGGCTCA CTGCAACCTT CACCTCAGCC
TTCAGGCTCA AGTGATCCTC CTACCTCAGC CTCCCAAGTA GCTGGGATCA      17200
CAGACGCACA CCTCTACACC CGGCTAATTT TTGTATTTTT GGTAGAGACA
GGCTGTTTCG CCATGTTGCC CAGGCTGGCC TCAAACTCCT GAGCTATGGC      17300
GTCCACCCAC CTCGGCCTCC CAAAGTGCTG TACTGGTGTG AGCCACTGCA
CTCCGCCGAG CCTTGGGCTT TCGGATGGGC TGTTTGCACT CGGGAACCTC      17400
AGAAGGGTGC TCCCCAAGAC AGAACATCAC CACTGCAAAC GGAGCCCACC
TTGCCAGGCA TTTCCCCAAC TGCCCTGATG GGAGGGGAAG GTGTGTCTTT      17500
CATGGCACAC AGGATGCTTC CCGCTGACAT GGACCCAGTG TCTTAAGCTC
TGTTTGCCCG CAATTCTTTT GTGTGGGGGA AGACAGGGAC CAGACTTAAT      17600
TTAATGTGAT TGTGTGGCGC CCTGGCAGTG GGATCAGTGT GGCTTCCTTT
ATGGAGTGCA GCCCTGCCAT AGCAACCTGC AGGCTGTCTT CTAGAGCCTA      17700
GAAGAGCTTT GTTCCGCAGA CATGACTTCC TTTTTTTTTT TAATGGGAAA
CGGGGAGGGC GTGGGGAGAA GGGCTCTGAT ATAAACAGTT ACCAAGGAGA      17800
CCCGAGTCGG GGTTGGAAAG GTTTTCTCCT TCCCTCGGTC CATACAGAGG
GAGGAGGGGT GTGTGCCCGG TGCAGGGACT CATGGGAGGC TTCACTCTGG      17900
AATCCTGAGT CCTGCCTGCT GGGGCTCCAG ATGTTGCAGA GTTCTTGGGC
```

FIGURE 1G

```
TCATGGGGGC AGACGCAGGC AGGAGCTCAG GGAAATTCCG AGAGGACTTT    18000
GGGGTGTCCA GGAGCACTTT GAGGCCACGG GCCCAGGCAA CTTCCTCACC
AGTACGGAAC AGGGGATGCA CAGGGCCGAA CTTTCCATCT GCTGGCCCCT    18100
GCTGGTGTCC AGGAGAGAGC TGTGACCCTC CTTTTTCCTT AAGTCCTGGA
GCTGAGAGTG AGAAGCCATG GGGTGCACGT ACTCGTCTCC ACCTGAGGAG    18200
CCCGCTCTGA GAAGGTCAGT TCCTGACACT GTCCCTGGCG GCAGACTGCC
CTCATGCCCC TTGGCCAGGG GACGCCTGCA CCCTCACTCG AGTGTGGGAG    18300
ACTTCCTCCT GGCGGGGACC TGCTGGAGGC CTGAAGATGC GGCTGCCTGG
AGTTAACAGC AAGGTCAGCC ATTCCCAGGG TTGCTGCCCC TGTCTCTGGC    18400
GAGGGACTCA GCCAGGCCTA GGCACTGGTG GCCCCTCTGG TCCAGGACTT
GCACGCTGCA TTTCCCTGCT TCCAAAGGAA GGAATTTGGA CTGGAGACCA    18500
GGAAGGCCTC ATGCACCCTC TTCTCTGCCT ATAAATAATC AGGTCCCCCA
TGCAAATCCA GGTCTGAGGT CCTCCGATGG AATGGAGCCC ATTCAGAGCA    18600
GTCCAGTGTG GGCCCTCTCG GGCTAAGCCC GTGAACTTTG CTTCTTGGAG
CCCCCAGCAG GGCCTTGGGT TGGGAAGGGA GTTTCCTGGC ACCCTCCCCG    18700
GCAGGATGGG TTTAACCAGC AGGAACTTGA AGGAATTGTT CTGTGTTGCG
CCGGTGACCT AGAGGCACAG GTCCGCCTCC AGCTCCTTGG GTTCCCTTTT    18800
CCCCCCTGGA CCACTCACAC TTCCTGGGTG AACCCTAGCC TTGGGGCAGA
GCAGGCTGGC ACACCTTGTC CACTCTGACG TTCCAGCCCA GAGGCTCCCC    18900
TGTCCACTCC CAGGGCCAGC CAGCCTGGCA TTCTGCTTTG TCCAACTGCG
TCAGGAGGTC ACATGGAAGC CCTCCTCCCC CGCCACTTGC CGGGCTTCTG    19000
GGATGGGAGC CGAGCTTGGA AGTTGCCATC CCTTGGCACC CCGCTTCCCG
TGGGGCTGTC TCTCCAGGGG GCCGTTTTCT ATTCCTGAGA AGACAGCCCA    19100
GCTCGTCCAA GAGGCCTTTC CCGGTTTCTT CAGCCATAAG CGTCTGCTCC
CTTTGAAGCC CTGTGGCGTG AACCTGTCCT GTCCCATCCG TCCCTGTGGC    19200
GTGAACTGTG CCATCCATCT ATGTGGTCTT AGCACATGCC ACCTTGTGCC
ATGGATGTCT TTGAAATCAC AGGCTGACCG CCTTCCAGGG AATGCCCTTT    19300
TTTCTACTGG GTGTTAACCA AGTTCCCCAG TCAGGCACTG GGGAGGCTGC
AATAACCAGA TGCAGCTCTT TAATAATAGT GTGGGCACGA GAGGGGCGGG    19400
GCCTGTGCAG TCCCGTGCCT GTCTAGTGTC AGAGCCAGGA TTTAAAGCCA
GGTCCGCCCT GCCACGTGGC TCTTCCCATC TTGCTGGGAT CCTCCTCGGA    19500
GGCTGTCTTC CCCTGGCTGG AGGATTGGGC CTGTGAAGTC ACAGGTTGTC
TCGTTCTAGT CTCATGCCCT CGGTGCCTGG CGCAGTGCCT CCCGTGCCCC    19600
CACCTGGCTG GCCTGTGTTA CCCGCCAGGC ACTCACGGGG CACTGGAGTT
TGCTCTTGTC CTTTTCTTCA GCAAATCTTC AGATCCACTT TGTTGGAGAT    19700
GGAACTTCCC TGGAAGCATC TGTCCTTTGA GTAGAGATAG GGGTTGGCTT
TTCAGCTGAT GCCAGGTCAG ACCTCAGAGG TGGCTAGGGA GTGAGAAGTG    19800
CCCTCAACAG CCTCTTAGTC CGCCCCTGCC AGCCTCGTGC TGGACAGAGA
ACTCAGGGCT GGCCACCCTA GGAGGAGACA CGGCGTCCTG TGCCGTGTGC    19900
TAGGCGTCGC GCTCATGGCA GGGATACACA GACAGCCCCC AAGGTTATTC
AACCCAGGCT TCCAGTTAAG CCACAGGAAT GTGCAGAGTA AAGTGCCTTG    20000
GCCGACCAGT GTGTGTGCCA CAGCGAAGAC GGCGACAGTA GCGTTGCTGC
GTGAGGGGCA GGGACGGCCT CTGGGAGAGG AAAGTTAAGT TCTGTGGTAA    20100
AGGATGAGTA GAAGGTGGGA AGGGGTCACA GGAAGGGCCT TGTAGGCCCA
GGGAGCCGCG TGGAAATGTG AGTAGCTTAG CGTACGCCAG AGAGGTGGGA    20200
AATGGGGCCG GGCCACTGCG GGAGGGGTGT TCAGAATCGG CCACTTTTCT
CTTGTTTAAA ATATCTGCTC TGCTGGGGTC CCCTTACTCA GGGTGTAGCT    20300
ATTCTAATTA TCCTTATTGT TGGAGGCTTC AAAATCAATC AGTCGGGCCC
CACATGTACC ACTTCTGTTT ATTTATATCG GTTGGTTGGT TGATTTACAG    20400
GGTCTCCCTC TGTCTCCCGG GCTGGAACCC GGGTGTTTGA AGCTGCAGTG
AGCTATGATC GCACCGCACG GTCATGTGAT CGCACAGTGG TGAGATCATC    20500
GCCCACTGCA GCTTCAAACA CCCAGGTTCA AGGGATCCTC CCACCTCAGC
```

FIGURE 1H

```
CTCCTGAGTA GCTGGGACCA CAGGCACAGA CCACCACACC GGGCTAATTA          20600
AAAAAAATTT TTTTTTTTAG AGGTAGAGTA TTGCTGTTTT ACCCAGGCTG
GTCTCAAACT GGCCTCAAGT GATCCTCCCA CCTTGGTTTC CCAAAGTGCT          20700
GGCATTACAA GTGTCGGCCA CTGAGCTAGG TTCTTGGATA CTCTCTTTAA
GAAAATGCCT CTTCTTGGAA GTTGCAGTGA GCCAAGATCG TGCCACTGCA          20800
CTCCAGCCTG GGCGACAAAG CGAGACTCCA TCTCAAAAAA AAAAAAGAAA
AAGAAAATGC CTCTTCAAGT CTTTTGATAG TTTTTCTGAT GGAGGATTTG          20900
TCTTTTTCCT GTTGATTTGT AGGGCTTTCT TACATTCTCG ATCAAAGTCT
TTTGCCCACA TATAAGTACA TTACAAATAT CCTCTCCCAC ACTGCAGCTT          21000
GCCTTTCCTT CCTTTAATGG CGTTTTTTAT TTTATTTTAT TTTATTTTAT
TTTATTTTAT TTTTTGAGAC AGATTCTCAC TCTGTTGCCC AGGCTGGAGT          21100
ACAGTGGCGT GATCTCGGCT CACTGCAACC TCCGCCTCCC TAGTTCAAGT
GATTCTCCTG CCTCAGCCTC CCGAGTAGGT GAGATTACAG GTGCGCGTCA          21200
CCAGGCCCGG CTAATTTTTG TATTTTTAGT AGAGACGGGG TTTTGCCATG
TTGGTCAGGC TGGGCTCAAA CTCCTGACTC AAGTGATCCT CCCGCCTCAA          21300
CCTCCCAAAG TGCTGGAATT ACGTGCATGA GCCACTATGC CCGACCAAAA
ATGTCATTTT AAATACAAAA ATGTATTTTC ATCTTCAGGT TTGTCACTTG          21400
ATTCAGAATG GCCATATTTC AAATCTCCAC CACCCAGAGG AAGGTGTTTC
CCACCTTGCT GTATTCAGCC AGCGTTTTTC ACCTCCTTCA GCTTGAAAGT          21500
CCTGGGCAAG GGGCACTGGG CTTGTTCCTC TGGCCCTGCA GGCTTGGGGC
AAGGGTCAGA AGGAAACAGA AGAACGGAAA ATGGGGCGGG GTGAGGGGTG          21600
GTGCGGGGGT AGAAGGCCAG AAGTGGACAG TGGACAGAAG ATTGTGTCAC
CCCAGCCTGA TCCTGAGGCT GGCGGGCACC AGGCTGCCCT CTGGAGGGGA          21700
TGGACCAAGG TGAGGGCTGT AGTCCCAGGA ACGCAGAGAA GGGAAAAATG
AAATGTGGCC CAGTGCAGGG GAAGCAGAGG GGGTCTTTGA ATCCCAGGGG          21800
AGCTGACCCT TGCGCAGGAC CTTGAGGGCA GAGCGGGGGT CAGCCAGGTG
AGGGGACTGT GGCAGGAAGG CTCTGGGTGG AAGCGATGGC GTGCACAGGA          21900
TTCTGAGGCC ACCATGCAGG GAGGGGGTGT GAATTAGGCA TCAAGCTTGG
TTTGTGGAGG GGAGTGGCAG GTGACGCTGG AGACAGACTC TGAGGGGACT          22000
CATGTGCCCG GATCCCATGG GTCGTGGGAG CTGTTGAAGA GCGGAGCAGG
GAAGTGGCAG ATGTGAGAGG CAGGGGCTGG AGAAGGAAGA GGTGAAGAGG          22100
TGGAAGGTCA GGCTGAGGTC AGGCCGTGGG GACCTGGGCT GTGGGTGGGG
AGGAAGAGGG AAGTGGAAAC CTTGGCTCTT CCTTGAGGGA AACAAGCCCA          22200
TATCTATCTG GGTGCAGGAG AGGTCTTGGC TTTGCAAAGC ACTGGTGTCG
CAAGAGGCTT TACTGTAGGG GAAATGGGAA GGGTGAGCTG GTGCGGGAGG          22300
CCTGGGAGCC CTGGTGTTGG ATTAGAGGCC AGCTGCTGGC TTAGCCCCTG
GGGAGTGGCA CGGGACAAAG CTGGAATGTC CCCTCCCTGT GCTGCCACAG          22400
CATCCCATGA AACCTGCATC TAAGGACACC CACGTCCCGT GACACCTGTG
TCCCATGACA CCTGCCTCCC ACGCTCATTG CACACTTTTG TCTTGCAGGA          22500
       [exon 3: 22499..
CCTCCTCTGT CAGAGAGAGC AGCTTTGTGG AAAAGATGAA GAAAACGGTG
            ..22547]
AGTTTCCTGC ATGTCTTTAC TCTTCTCAGG AAGCCTCGGC CCCGATGGGC          22600
ATGTAATCAA GTCTAGCAGA CGGCTGGCGT CACCGCATAG AGGATGTCCC
                                        A
GGTGGCCAGG AGAGGGAACG CAGCCCGGCT CGCTTGGGAG TCATTCTTGC          22700
CTAACCAAAG CCCTGAAGCA GCTCTGCTGT TGCTGGACA GTTTGTTTCC
TTGAATCTTT ACTTTTTGAA CAAACCAAAG TCTAAGAACA GACCCTCCCT          22800
CTAACTCACG AAGCTTCTCC CGTCTGGCCC TGGCCTGTCC AGGAGCAGAG
GTAAATCTCA CAGAGCTGCG GTCGTGGCAC CACTGCCATT TTCTCTCCGC          22900
TTTGTTTTTG TTACTTTTTT TTTTTTGAGA TGGAATCTCG CTCTGTCACC
```

FIGURE 1I

```
CAGGCTGGAG TGCAGTGGCG CAATCTCGGT TCACTGCAAG CTCCACCTCC          23000
CGGGTTCAAG CCATTCTCGT GCCTCAGCCT CCCGAGTAGC TGGGATTACA
GGTGCGTGCC ACCACGCCCA GCTAATTTTT GTATTTTTAA TAGAGATGGG          23100
ATTTCACCAT GTTGGCCAGA CTGGTCTTGA ACTCCTGACC TCAGGTGATC
CGCCCGCCTC AGCCTCCCAA AGTGCTGGGA TTACAGGCGT GAACCACCGC          23200
ACCCAGCCAA TTTAACTATT TAATTGTAGG CATCAGTACT ACATTTACAT
AGTTAAAATA AAAACATCCA CTATTCAGTG GAAAGCTCCC CTCTGCCACC          23300
CACTCATCCC CACTGTTCTT ATTATTTATG GATCCTTCTA GAGTTCCTGT
ATGAACATAC AAGTAAAACA AACAGAGACT CTTATTTTTA ATTTAAAAAA          23400
ACCTTTATTT TATTTTAGAG ACAGGGTCTC ACTGTGTTGC CCAGGCTGGA
GTGCAGCGGC GCCATCATAG CTCACTGCAG CCTCAAACTC CTAGACTCAA          23500
GCCATCCTCC TACCACAGCC TCCCAAGTAG CTGGGACTAC AGGTGCACAC
CACCACGCCT GGCTAATTTT TTAATTTATT CTTCATAGGA ACAGGTTCTC          23600
GCTATGTTGC CCAGGCTGGT CTTGAACTCC TGGCCTCAAG CAATCCTTCT
GCCTCCGCCT CCCAAAGTGC TGGGACTGTA GGTGCGAGCC ATTGCATCCA          23700
GCCTACACTT TTCTTAAAAC ACAAAACGTG GCATACTATG TGCACTGTTC
CAGACCTTGC TTTTTTTTTT TTTTTTTTTT TTTTTGAGAC AGAGTCTCGC          23800
TCTGTCGCCC AGGCTGGAGT GCAATGGCAC GGTCTTGGCT CACTGCAACC
TTCACCTCCT CGGTTCAAGT GATTCTCCCG CATCAGCCTC CGGAGTAGTT          23900
GGGGTTAGAC GCCTGCCACC GTGCCTGGCT AATTTTTGTA TTTTTAGTAG
AGACGGGGTT TCACCATGTT GACCAGGCTG GTCTCAAACT GCTGACCTCG          24000
TGATCCATCT GCCTCGGTCT CCCAAAGTGC TGGGATTACA GGCGTGAGCC
ACTGCGCCCG GCTGACCTTG CTTTAAAAAA AAAGAACAC AAAACAACTT           24100
GATATGTCTC AAAGATCCTT CCATATCAAT ACAAAGATAC TGCCATCATC
TTAGCTGTAC CATGATTCCA TTATATGGCC ATCTGGATTC ACTGTAACGC          24200
ACTTAACCCT GTCCTATTAA AGATGGGCAC TGGGGCTGTT GCTACTCTTT
GGCTCTTACC AACAATGTCA CTGTGAATAT GTAGCTCGGT TTGCACATTT          24300
CAAATACAAA GCAGTCCCAG AAGTGGAATT GCTGGATTGG AAGGTAGATG
TGTTTGCAAT CGTGATAGAA GCTACAAATC GTCCCCGTAG GAGTCACACC          24400
GTGGCACGCG CTCAAGAATA GCGACGTGTG AGAAAACTGT TCCCCACAGC
CTTGCCAACA GTTTGTTATT AAAGTTTAGA TTTTTGTCAG TTTTCTAGGC          24500
TTAAAAAAAA AACTCAGTAT AGTTGTAATT TACATTTTTA TGAATGAGAT
TCAGTAGATT TTCACCCACC TAGGTCTAGA TGGCTTCGTT TCTTTTTCTT          24600
TTTTCTTTTC TTTTTTTTTT TTTTTTTTTT GAGACAGGGT CTCGCTTTGT
TGCCCAGGCT GGAGTGCAGT GGTGTGATCT CAGCTCACTG CAATCTCCAC          24700
CTCCCGGGTT CAAGCAATTC TCATGCCTCA TTCTCCTGAG TAGCTGGGAT
TACAGGTGCG TGCTACCATG CCTGGCTAAC TTTTGTATTT TCAGTAGAGA          24800
CAGGGTTTTG CCATGTTGGC CAGGCTGGTC TCGAACTCCT GACCTCAGGT
CATCCACTCA ACTTGGCCTC CCAAAGTTCT GGAATTACAG GCCTGAACCA          24900
CCGCACCCAG CCTCTTTTTT CTTTTTTTAA TAGATTCTCT CTTCACATCT
TTTGCCTGCT TTTCTGTTGG GATTGATCTT TCTCCTTGAA ATTATTAGGG          25000
AGATGAATCT CTTGTCTGTG ATATAAGTTG AAAATATCTT GCCCGTTAAA
GAGTTTGTCT TTTAGTGAGA TTCCATCTCT TAAAAAAAAG TTTGTCTTTC          25100
AACCTAGCTG ATCGCATTTT TATTTGTTTT GCTGTGTAAA AGTTACAGAG
TCACCTATAT GATTCTTTCT TTTCGTGACT TCTGGTCTTA GAGTCATAAT          25200
TTGAAGGTCT TCCCCATGCT AAAGAGATAA AGATATTCCT GTATTTTTTT
TTTCTAGTAC TTTTAGGGTT TTGTTTTTCA TATTCAAAAT TTAAAATCTG          25300
ATCTATTGGG AATTTATCCT GATGTAAGAT GTGAGGTATG AATCCAACTC
TACATTTTAC ATAGCTGTCT AAACACCATG TATTAGAGTC CATCTTTCCC          25400
CCACTGGTTT GAAATGCCAC CTTTATCATA TACTTAATTC CTGGGCTGGG
CGCAGTGGCT CATGCCTGTA ATCCCAACAC TTTGGGAGGC CGAGATGGGC          25500
AGGTCGCTTG AGCCCAGGAG TTTGAGACCA GCCTGGGCAA CATGGCGAAA
```

```
CCCCATCTCC ACAAAAAATA GAAAAATTAG TGGGACACAG TGGTGCGCAC     25600
CTATAGTCCC CGCTACCTGA GAGGCTGAAG TGGGAGGATA GCTTGAGCTT
GGGAGGCACA GATTGCAGTG ACCACTGGAC TCAAACCTGG GCAACAGCTA     25700
CACCCTGTGT CAAATAATAA TAATAATAGC TGGGTCTCTG GGTTCTTTTC
ACTGGGCGCA ACGTGTTCAA GGTTCACCCA CACTGTAGCA TATGTTAGTA     25800
CTTCCTTTTT ATTGCCAAAT AATGTTTTAA TATATGGATT CTTACTTATT
TTTCCGTGAC TTTTAGGATC AATTTGTTGA GGAAAAAAAA GCTGTTCAGA     25900
TTTTTTTTTT TTTTTTTGGT AGAGACTGGG TCTCACTCTG TTGCCCAGGC
TGTTCTCAAA CTCCTGGGCT TAAGCAGTCC TCCTGGTGTG GCCTCCCAGG     26000
GTACTGGGAT TACAGGTGTG AGCCACTGCA CCCAGCTGGT TCAGATTTCT
ATTACATCAT GCCACATTTC TCTAGTGACA TAAATAGAGC TGCCATTTTT     26100
GTGGGTCGAG ACTCCCTGTC CATAATCGGG GTGCTTTCC ATTGGCGTGA
GTCCTTTGTG TGTAGTTCAG TAGCTTCTGA CACATCTCTT ACAAAGTTTA     26200
TTTCTAGGTA TTTTATATAT TTTGATGCTC CTGTAAATAT TTAGCTGCTA
TCTTAGGTCA TCTTTTGTTC TCTTAACCTG ATAATCATGA GCACTTTCTC     26300
TTCTATTAAG TATTTTTACG GTGATTATTT TAAAATATCC TTTTGGGCGA
AAACCCTGTC TCTACTAAAA ATACAAAAAT TAGCCGGGCG TGGTGGCATG     26400
CACCTGTAGT CCCAGGTACT CGGGAGGCTG ATGCAGGAGA ATCACTTGAA
CCCGGGAGGC GGAGGTTGCA TTGAGCCAAG ATCGCACCAC TGCACTCCAG     26500
CCTGGCCAAC AGAGCAAGAT TCCATCTAAA AAAAAAAAAA ATCCTTTTGG
CCGGGTGCAG TGGCTCATGT CTGTAATCCC AGCACTTTGG GAGGCTGAGA     26600
CGGGCGGATC ACCTGAGGTC AGGAGCTTTG AGACCAGCTG GACCAACATG
GTGAAACCCT GTCTCTACTA AAAACACAAA AAATTAGCCA GGCATGGTGA     26700
CGGACACCTG TAATCCCAGC TATTCAGGAG GCTGAGGCAG GAAAATCGCT
TGAACCCAGG AGGCAGAGGT TTCAGTGAGC CGAGATCGCA CCACTGCACT     26800
CCAGCCTAGG CGACAAGAGC AAAACTCTGT CTCAAGAAAA AAAAAAACAA
TTTTTAATGA TTTAACAATG ATTTGTGGAA AATTCAGGAA CTTTAAAAAG     26900
TATAAGAACA TTTATTGTAA TCCCTAAGTT TTTTTTGGTG GTGGTTTTGT
TTTGAGACTC AGTCTCACTC CTGTCACCCA GGCTGGAGTG CAGTGGCTCA     27000
ATCACAGCTC ACTTCAGCCT CAACCTCCCA GGCTCAAGTG ATCCTCCCAC
CTCAGCCTCC CAAGTAGCTA GAACCATAGG CATGCGCCAC CACACCCAGA     27100
TAAATTTTTT GTATTTTTAG TAGAGACAGG GTTTTGCCAT GTTGCCCAGG
CTGGTCTCAA ACTCCTGGGC TCAAGTCATC CTCTTACCTC GGCCTCCCAA     27200
AGTACTAGGA TTACAGAAGT GAGCCACGGC ACCCGGCCGC GTAAGTTTTA
TTTTGTATGA GGTGGGATTC AGTGTGAAAA CTAGAGACTA CTTTAGGCAT     27300
TTTAATCTGG AAGGGATTTC ACACAGGGAA GGGGGTACTT CCAGGTTACT
GGAGGGGCTG AAGGAGCAGG TTGGTCTCTA AGATATCACA GAAGCAACCT     27400
GCAAGGTAGC TGCTCTGTTC GAATGCAGCC TGAAACTGTT GAGTTCAGGA
CACACCACGT GTAGCTGTGA TCCAGGATAA GAAGCTACTT CTACAGCTGC     27500
CTCTGAACAT CACATAAGCC AGTAGCCAGG ACTAGGAACT TGCAGTCTGG
CTACCAAAAC TATTAGAAAT ACCTCAGGTG GCTGGGTGCG GTGGCTCACG     27600
CCCGTAATCC CAGCACTTTG GGAGGCCAAG GTGGGCAGAT CACCTGAGGT
CAGGAGTTCG AGACCAGCCT GGCCAACGTG TTGAAACCCA GTCTCTACTA     27700
AAAATTCAAA AAAATTAGCC TAGCATAGTG GTGGGTGCCT GTAATCCCAG
CTACTCAGGA GGCTGAGGCA CGAGAATTGC TTGAACCCAG GAGGCAGAGG     27800
TTGCAGTGAA CTGAGATCGT GCCATTACAC TCCAATCTGG GCGTCAGAGC
GAGACTCCGT CTCACACACA AAAATAAAAA AAGAAAAGA AAAAAGAAA      27900
AATGCCTTAG GCACTGAAGA CCAGGCACTG CAGAGAACCT CACATGTGCA
TGAGGGGCCA GGAAGCCTTG GCCTCCCTTC CACCCTTCAG ATCTTGCAAA     28000
GCACGTTCAA TTGGCTGAAC CGAATTGGCT GCAACTCCAG CCCCAACGGG
CTCTGGGGAA TAGGATTTCA GGCTTTTCAA CAGCTGCTGG AGGAGGAGGG     28100
TGGACTGACG AGTATGTGAG CCAGCCCCTA GGACCTTTCC CTGTTGTAAA
```

FIGURE 1K

```
ATTTTATAAT GAATATGTAT TTAACCAGCA AAACACATCC AGAATGCCCC      28200
ATTTAGGGAC CCTCGGTATA TTTCTTGCCA GTCTGTTTTC TTTGTGCCTG
TGGGTATAGA AAGAAAATAC ATATGTATGT TTTGATATAC ATATCTATTA      28300
TTTATATAAC ACAATTGGAT TGTATAGCCA GCCCTGCATA TCCGTGAGTT
CGGAGTTCAA GAATTCAACT GCAGACTGAA AATATTTGGA GAAAAAATGG      28400
ATTGTTGTGT CTGTCCTGAA CACATACTGT CATCTTACTG TTCCCTAAAC
AATACAGCAT AGCAACTATT TACAGAGAAT TTACATTACA TTAAGTATTA      28500
CATTAATAAG TAATCTAGAG ATGATTTAAA GTATGTTGGA GGATGTGTGT
AGGCTACACA CAAATACCAC ATCATTTTAT AGCAGGGACT TGAGCATCCT      28600
TGGATTTTTG TTTTTTGGGT TTTTTTGTGA TGGAGTTTTG CTCTTGTAGC
CCAGGCTGGA GTGCAATGGC GCGATCTCGG CTCACCACAA CCTCTGCCTC      28700
CCAGGTCCAA GCGATTCTCC TGCCTCAGCC TCCCAAGTAG CTGGGATTAC
AGGCATGTGC CACCATGCCC GGCTAATTTT GTATTTTTAG TAGAGACAGG      28800
GTTTTCTCCA TATTGGTCAG GCTGGTCTTG AACTCTCGAC CTCAAGTCAT
CCACCCGCCT TGGCCTCCCA AAGTGCTGGG ATTATAGGCG TGAGCCACTG      28900
GAGGTGGAGT CTTGCTCTGT CACCCAGGCT GGAGTGCAGT GGCGTGTTCT
CAGTTCACTG CAACCTCCAC CTCCTGGGTT CAAGCTGTTT TCCTGCCTCA      29000
GCCTCCCTAG TAACTGGGAT TACAGGTGCA CGCCACCATG CCCGGCTAAT
TTTTGTAGTT TTAGTAGAGA TGGAGTTTCA CCATGTTGGT CAGGCTGGTC      29100
TTGAACTCCT GACCTCAGGT AATCTGCCAG CCTCAGCCTC CCAAAGTGCT
GGGATCACAG GTGTGAGCCA CCACGCCCGG CTGCATCCTT GGATGTAAGT      29200
ATCTATGGGA GGTCCTGGAA CCATTCCCCC ACGGATAGGG GCTATACATT
TGGTTACTTT TTTTAACTTT GTTTTTTCCC CGTGTCATGA AAATGTCAGA      29300
AGTGTGATCT TGAGCAGCCG CGTGTGTGAG ATTGCCTTGG TGACCTTTGC
CCTCCTTTGC CACAGTGGCT GTGACAGTGA GAAGCAAGTC CCAGAGGAAC      29400
TTAGAAGGGA CTCAAAGCCA GGAAGGAAAG GGGGCGGCCT GGAGGGCCCC
CGCCTGCCAG GCCTGCCCAG TGGGTGTTCA CCGGAGCCGT GGCTGAGGTC     .29500
TGTGGCCCTC ACCAACCCTG TGTCTGCCTT CCTTAGGGGA GGAACATCAT
      [exon 4: 29537..
CGTGTTCTAC GGCTCCCAGA CGGGGACTGC AGAGGAGTTT GCCAACCGCC      29600
                          T
TGTCCAAGGA CGCCCACCGC TACGGGATGC GAGGCATGTC AGCGGACCCT
GAGGAGTATG ACCTGGTAAG CTGCCACCGC GTGCTGGCCC CAGATGGAGG      29700
              ..29665]
CAGTGGGTAG GACAGGGAGC AGGTCTGTAG GGCGCCCCTC AGCAGGGGGA
GGCCGGCAGG GAGTGGGGTC CTGGGAAGAC GTCCTCGGAA GTTGCCTTCC      29800
  T
CGTGAGGGTC ATTGCCATCC CTGCCCGGGC TGACTGACGG AAGGGCCATT
CCTGCAGTTG CAGCCACGTG CCAGGCATCG TTTCCCCAGG ACTGGCCCTC      29900
CCCTGAGTGC CGGCTGCCCT CCTGGCGGCC GCCACATCTC CCAAACCAAA
CCCACCCTCC CCAGCTGCTC CACTCCCCTC TCCTTCGAGG TCTTGGTGAC      30000
GGGCACTACC CCACAGCCCT CTCCAGGCTG CATTTGCAGT TGATGAACAG
TTTAGTTAGA TATCAGTGTG CACCGGGGCT GTGCGGTGCC TGGGAGCTTT      30100
GGAATGAGGT TCCCAGGGTG GCCCCAGACC TGCATCAGCT CAGGAGGCTG
CATTCTGATG GCTCTTTTGT GCCAAGCTTC ACTTTCCTCT AGAAACACAT      30200
CCTCTCCGCT CCAGCCTAAC ACGGGTGACC TTGTCCTGCC CACCTGGCAC
AAATGCCCAT GCCCCAGCCC CTCCGTGTTG TTACTTCTCT CTGATCCCAC      30300
GACACTCAGA CATCCCTGGC CTGGTGCCAC CCTGGGCAGG ACCTGGCCTT
CCCCATCTGG TGCGGGTTGA ACCTTGAACA GGCTCAGTCA TGGCCGGGGC      30400
GCGGTCCTGT CCCTGTTTCT GCAGGCCGAC CTGAGCAGCC TGCCAGAGAT
                          T
```

```
[exon 5: 30425..
CGACAACGCC CTGGTGGTTT TCTGCATGGC CACCTACGGT GAGGGAGACC         30500
CCACCGACAA TGCCCAGGAC TTCTACGACT GGCTGCAGGA GACAGACGTG
GATCTCTCTG GGGTCAAGTT CGCGGTGAGT CACCCAGAGA CTGCTATGGG         30600
              ..30574]
CTCCCGGTGG CCTGCGGTGC CTCCCTGGGG ACTCCAGATC CATGTATCTG
AAAGGCAGCC CTCCAGACCC CCACCCTGTC CTCAGCAGCC AGGGCAGGCC         30700
ACTTGTGAGA CCCTCTCCTC TGCCCCAGAC CCCAGCACTA CGAGAATGTC
CCCTCCCTGT CCCCAGCCCC AGTCGGCTCC AGGGGGCATT GGGTGCTGGA         30800
GACTAAAGGC AAGAAAGGTC TGGCTGGACG ACTGAGACCT GCCTGGCCTC
GGAGAGCTGG AGCCCCAGCC CGGTGGGGAG GGTGGGCTGG CCCCGCTTCC         30900
CTGTGGGTTG AGGCTGGCAG TGGACGGGGC AGGCTGTGGG CTGCAGGTCA
ACCAGATGAA GCCTCTTCAG TGGCCCAGTG TTCCTTGCAG TGCGAGGCGC         31000
CTGGTGGAAC GGAGGCCTGC AGGTGTTGCC AGCAGCTCAG CCAGTGCTGT
GGGGCCTCCC GCCCTGCCCC CATGGCCCCT CCCACTGGTC AGGTCGAGGG         31100
CCAGGCCTCA GAGCGGCCCC TGTGTCCACG CAGGTGTTTG GTCTTGGGAA
              [exon 6: 31134..
CAAGACCTAC GAGCACTTCA ATGCCATGGG CAAGTACGTG GACAAGCGGC         31200
TGGAGCAGCT CGGCGCCCAG CGCATCTTTG AGCTGGGGTT GGGCGACGAC
GATGGGAAGT GAGTGCCCAC CCTGCCACCA TGATCAGCGC GGCGGGCTTA         31300
              ..31258]
GGCAGGGGCC GTGGAACGTG AGGGGCGCGC ACACCATTGT GTCAGCTGAG
ACTCAGCGAC ACGCACCTCC AACACAGAGG GAAGGGGCTC TCCTGCCTCT         31400
GCCCTGCCCC TGCCAGTTTT GCTTTTCGGC TTGCCCAACT CCCTGGAGCC
TTCCTGATGC TCTGGGTTTA TGTCGCTGGG TGCCCCAGGG TGCACAGTCC         31500
TGAGCTTTGG GGATGGGGTG GGGTCGGGGC GTGCCTGGCA CCAGGTACCG
                                    A
TTGCCACATG GGCCTCCCCT GAGCCGCTCC CCCTCTCCTC TCCTCGGCCC         31600
AGCTTGGAGG AGGACTTCAT CACCTGGCGA GAGCAGTTCT GGCCGGCCGT
                                             T    T
              [exon 7: 31603..
GTGTGAACAC TTTGGGGTGG AAGCCACTGG CGAGGAGTCC AGGTGAGCAA         31700
              ..31692]
GTGCCCGCAG GTGCGGTGGG TGGCCTGGGC GGGTCCTGTG CCGAGGGCAG
CCACCCTGGA ACAAGGGCTG GCAGTGGGTC GCAGCAAGGT TAGAAGACAC         31800
TCCGTCATAG GGTCGAGGAG GGACCTTGGT CCCAGCCAAG GACTCACTCT
GCCACGTTGC TCTGCACTGC CCTGGGGCAG CGGGGTGCAT CCCACCTCTC         31900
GACAAGGACA CATCGCGTCG GGCTCTGTGG CTAGGTTCAA CTTTGGCAAA
AGGGCTCATT TCCTTAAAAT CTGCCTCCAC AGACTTGGCC AAAAACATAA         32000
CGTTCCTCTC TGTGCCCTTG ATGGCCCCTG GGTGCTGCCC GGGCTTCCTT
ACCTTCTCCC AGATGGAAGC CTGCCCAGCC CTGCCCCGGC TTCTGGGCGT         32100
CTGGCCCCCG GCAGCTCCAC GCCGCCTCCC TCCTTGCTCC CTCGCCTGCC
CTCCTTGTGC ATCTGCAGCA GGGGCTCCCC TGCTTCTTGT CGTATGTACC         32200
TGGGACCTCA CCCCAAAGGC CATGCACGGT CTCCCCTGTA GTCCAACCCC
TCCCTCTCGG GACTGACCCC TGCCGCTTCC CGGCCTCACC CTTGGTCTCC         32300
CCTTTCCAGC ATTCGCCAGT ACGAGCTTGT GGTCCACACC GACATAGATG
              [exon 8: 32310..
CGGCCAAGGT GTACATGGGG GAGATGGGCC GGCTGAAGAG CTACGAGAAC         32400
CAGAAGCCGT GAGTGGAGGG AGCGTGGCTT GGGGCAGACG GCTCTATGGC
                                             G
    ..32408]
```

14/19

```
CACTGGTGCA CCCCAGGCTC AGTCTGCCGT GTATCCCCAT ATCCCCACAG        32500
                             T
GGCCCTTCTC ACCAGACCCC GTGCCCCGAG TGGGTGTGAG TGTCCACGAC
CTGTCCACGG CCCGGCCCCA GGAGACTGGC ACTCGCCCAG TTTGTACTGA        32600
GCTCACTTAG CCACCTTACT CTGCACTGCC CAGGGCAGCG GGGCACATCC
CACTTCTCGA CTAGGACAGT GCATTGGCCT CTGTGGCTAG GATCCTCTTT        32700
GGCAGAAGGG CTCATTTTCT TAGAATCTAT CTCCATAGAT AGACTTGGGG
GAGAAACCTG GCGTTCCTCT CTGTGCCCTT GATGGCCCCT GGTTGCTGCC        32800
CAGATTCCTG CCCGAGCCAC CCAGGGAAGT CCTCGTGCCA CTGCATCCAA
GACCAGCTGC CCTGCTTTCT GTAGGCCTGC CCCAGCCTGG CTGGAGCGAG        32900
AAGCCCTGCA TGTCGGTGGC TTTCCGGAGT GCAAGTTTCT CTCTTGTCCA
CGGTGTAGTC CAGGACGTGT GGGGGTGGGG GAGCTGTGCT CCTCATTTGG        33000
GGACCAGCCA CCTTCCAGCC TGTGGCACCG TCAGCTTGGG CCTCACAGTT
CTTTTAGGGG CCAGCCTCAG TTTCCACATC TGTAAACAGA CACTGATGAC        33100
CCAGCTCTGC CCATGTGGTG GGACTATAGA GAGAACTGCA TTAGGGCTGG
GCGAGGTGGC TCACAGCTCT AATCCCAGCA CTCTGGGAGG CCGAGGTGGG        33200
TGAATCACCT GAGGTCAGGA GTTCAAGACC AGCCTGGCCA ACATGGTGAA
ACCCCGTCTC TAGTAAAAAT ATAAAAGTTA ACCAGGCATG GTGGCGAATG        33300
CCTGTAATTC CAGCTGCTCA GGAGGCTGAG GCAGGAGAAT TGCTTGAACC
TGGGAGGCAG AGGTTGCAGT GAGCCGAGAT CGCGCCGCTG CACTCCAGCC        33400
TGGGTGGCAG AGCGAAACTC TGTCTCAAAA AAAAAAAAGA GAACTGCATT
GGACCAGGCT GGGAGAGCCC TTGATGTAAC CGGTGAGATT TCCTCATGGA        33500
GATCTCTGAG ATTCCCTGTG CTTTGTGCAA CCAGAAGCGT CCTTGGAGAC
                                                    G
GGAGACTCAG ATCAAAGCCC CGGCCGCTCA CTGTGCTTCT CTCCTCCCCA        33600
A                     T
CCCAGCCCCT TTGATGCCAA GAATCCGTTC CTGGCTGCAG TCACCACCAA
         [exon 9: 33606..
CCGGAAGCTG AACCAGGGAA CCGAGCGCCA CCTCATGCAC CTGGAATTGG        33700
ACATCTCGGA CTCCAAAATC AGGTACCAGC TGCCACTGTC ACCCCCTGAA
                   ..33722]
CCCTCACTCT GGGCCTCCTG ACCTGGGGCA GGGCCAGCCT TCCGCCCCTC        33800
CCGAGCCTCA CATCTCCCTC CAGGTATGAA TCTGGGGACC ACGTGGCTGT
         [exon 10: 33824..
GTACCCAGCC AACGACTCTG CTCTCGTCAA CCAGCTGGGC AAAATCCTGG        33900
GTGCCGACCT GGACGTCGTC ATGTCCCTGA ACAACCTGGA TGGTGAGTGC
                  ..33942]
CACAGTCAGG GCGCCCTGCC GGGCTCAGGC AGCCGCGGGA TTGGGCCTGT        34000
AGGAAGGCCC TGGGTTGAGC TTCTGCTTAG GCCTGAAGCC CCGGTGCCTG
GGAGGCCCTT GCACCGAGAC TCCACGGTTA CAGGATCCCA AGCAAACGGG        34100
AGGCGGGGTG GCCCTAGGGG TCTAGCCCTC TCTGTCGGGG TTCCCCCTAC
CCCGTCACTG TCATAGTCCT TTAAGGGAGT GAGGTGCTGA GGCCTGGTGG        34200
CAGAGGCAGC CCTGGCTCCC CCATGGCCAC TGTGTCCTGC TGGGAAGGAG
GGCCTGGCTC CACGACCCAC CTCTGCCGGC CTGGGGCTGC CCCCACCTCC        34300
TCACTGAAGT CAGGAGTCAG CAGCCCTCCC AGGCCCCAA GGGTGCACAG
TGGTGCCGAG TGGCAGTAGC CATTACGCGG GGCTGCCTGG GCCTGGTGGG        34400
GCTGCCCAGC CTGAGCCTCC CGCTGTGAAG CCCTCGGACC CCACTGGTCA
CCAACCTGGG CCGAGCCCAC CTCGCCCCAC CCCTGCTTGC CGGTCCTCAG        34500
CTGCCATGCC AGGGCTGCCC TTACTGTCGG CTTCCACAGC CCGCCTGTAG
GGAAGCCTGG GCGGGGCTGT GTCAGACCGT GTAGTGTGCG GTGAAAGCAC        34600
AGCGGGTGCG TTAGTGTGCA GGGGCTCCCC CGCACCTCCC CTAGCAGGGT
```

FIGURE 1N

```
CCTGCCTCTG ATGAGGACTT CCTGTCTGGT TGGAGGCCCC AGACTTGGCC        34700
CCAGGGCCAG GGAGGCATCA GAGAGCATAG GCCTTGTTTC CAGCACCAGC
TGGGGCACCT GTTGCCGCAG AGCTGGCCCA AGGTGTCACC CCCTCCTGCC        34800
           T                                        TC
GCAGCCACCC ATCCCCAGGA GGCGGCCGCC TACCCCAAGT CCTGCCTGTC
                                                      G
TCTTCCCTGC AGAGGAGTCC AACAAGAAGC ACCCATTCCC GTGCCCTACG        34900
     [exon 11: 34863..
TCCTACCGCA CGGCCCTCAC CTACTACCTG GACATCACCA ACCCGCCGCG
TACCAACGTG CTGTACGAGC TGGCGCAGTA CGCCTCGGAG CCCTCGGAGC        35000
AGGAGCTGCT GCGCAAGATG GCCTCCTCCT CCGGCGAGGG CAAGGTGCGC
                ..35044]
CCCCTCAGCC CCCGCAACCT CCGCCCCGTC ACCCCGCCGT TTTCCGAGCT        35100
CCGTGGGCCC CAATCAGCCC CATCTCACCC CCGTGTCTCT TAGGAGCTGT
     [exon 12: 35144..
ACCTGAGCTG GGTGGTGGAG GCCCGGAGGC ACATCCTGGC CATCCTGCAG        35200
GACTGCCCGT CCCTGCGGCC CCCCATCGAC CACCTGTGTG AGCTGCTGCC
GCGCCTGCAG GCCCGCTACT ACTCCATCGC CTCATCCTCC AAGGTGAGGG        35300
                ..35293]
CCGGCACTGC CCTGCCAGCC ACACGCTGGA GGCCCAGCCC TGCTCACAGC
                     A
AGGCAGAGTG CAAGGCGGCA CAGGAGCTCC GAGATCTGAG CCCTGAGCTC        35400
CAGTTCCAGC CCCAGCGCAG CTCCAAATGC CTCCCCAGGC TGTGGACTCA
GTCGGGCTGG CTTGTGAGAT TCTCAGCATC TGTCCAGCCC CGGTCCCCAG        35500
AACCAGTCCG GGAAGCCGCT GGGGAGGGGG CCTCTGAGGT TTGGGTGCCA
GGTGGGCTGG AAGAGGCCCT GGGTGAGTGG GGCTGGCCTG CAGAACGGGA        35600
CTTGGGGCCG GGGCTGGGCA AGGGCCTCGG TGTGGCGGTG GAGCTCACAC
                                            CT
GGCCCTCCCC ACAGGTCCAC CCCAACTCTG TGCACATCTG TGCGGTGGTT        35700
     [exon 13: 35665..
GTGGAGTACG AGACCAAGGC CGGCCGCATC AACAAGGGCG TGGCCACCAA
                     T
CTGGCTGCGG GCCAAGGAGC CTGTCGGGGA GAACGGCGGC CGTGCGCTGG        35800
                     C
TGCCCATGTT CGTGCGCAAG TCCCAGTTAC GCCTGCCCTT CAAGGCCACC
ACGCCTGTCA TCATGGTGGG CCCCGGCACC GGGTGGCACC CTTTCATAGG        35900
CTTCATCCAG GAGCGGGCCT GGCTGCGACA GCAGGGTGAG TGGGGTCCCA
                ..35935]
TGGGGGAGAG GGGGTGACGA CTGGGAGCCC CGCGCTCACC CCGGCCCCTG        36000
                                            T
CCACGCAGGC AAGGAGGTGG GGGAGACGCT GCTGTACTAC GGCTGCCGCC
     [exon 14: 36009..
GCTCGGATGA GGACTACCTG TACCGGGAGG AGCTGGCGCA GTTCCACAGG        36100
  G
GACGGTGCGC TCACCCAGCT CAACGTGGCC TTCTCCCGGG AGCAGTCCCA
CAAGGTGAGA CGGGCGGGCA CCCACGAAGG TGGGCATGAG GCTGGCAGGG        36200
           ..36154]
CCACAGCCAC AGTGCCCCCC TCACAGCACC ACCCTTGGCC CCAGGTCTAC
  G
```

FIGURE 10

```
         [exon 15: 36245..
GTCCAGCACC TGCTAAAGCA AGACCGAGAG CACCTGTGGA AGTTGATCGA        36300
AGGCGGTGCC CACATCTACG TCTGTGGGTG AGTGAGTGGG GTCACTGGAA
                            A
               ..36327]
TAGGGGGCAG GGAGGACAAG GCCCTGCCTG CCACAGTTGG CCCAGCCCCC        36400
AGCACCCCCT CTTCCTGCCC AGGGATGCAC GGAACATGGC CAGGGATGTG
         [exon 16: 36423..
CAGAACACCT TCTACGACAT CGTGGCTGAG CTCGGGGCCA TGGAGCACGC       .36500
GCAGGCGGTG GACTACATCA AGAAACTGAT GACCAAGGGC CGCTACTCCC
TGGACGTGTG GAGCTAGGGG CCTGCCTGCC CCACCCACCC CACAGACTCC        36600
               ..36567]
GGCCTGTAAT CAGCTCTCCT GGCTCCCTCC CGTAGTCTCC TGGGTGTGTT
              C
TGGCTTGGCC TTGGCATGGG CGCAGGCCCA GTGACAAAGA CTCCTCTGGG ·       36700
CCTGGGGTGC ATCCTCCTCA GCCCCCAGGC CAGGTGAGGT CCACCGGCCC
CTGGCAGCAC AGCCCAGGGC CTGCATGGGG GCACCGGGCT CCATGCCTCT         36800
GGAGCCTCTG GCCCTCGGTG GCTGCACAGA AGGGCTCTTT CTCTCTGCTG
AGCTGGCCCA GCCCCTCCAC GTGATTTCCA GTGAGTGTAA ATAATTTTAA         36900
ATAACCTCTG GCCCTTGGAA TAAAGTTCTG TTTTCTGTAT TTGCCTGGTA
TTGTGTGAGT AGATCTGGGA CCTCCACCTG CATCAACTTA ACTAACTCAG         37000
ACCCCAGGAA CCCATGTGGT GGGGCCACCC AGCCCTCCCC TCCCCCAGGA
GAACACACAC GCTCAGGCCA CCTCTGGGCC TCTCTTTATT GAGGGCACTG         37100
GGCCCAGGTC TTCCTTCAGG GCCCACAGCG CCCATAAAAC CCAAGGGAGA
ATAGAAGAGA CCCCCTGATA CACGCACACT CGAGGGGCGC CTCCCATCCC         37200
CTCCCACAAC ACACAGGACA GAAGCCCCTC TGGGCCGGCA GGGGAAGGCC
CAGCCTCAAT CCTTCTTGCT CCCGTGCCGC TGACTGTGAA ACTTGTGGTG         37300
CACAACCCTC AGGGTGGTGA AGAAATTGCC GAGGAAAAGG AGGAGGAAGG ·
GAAAGCCGCA CATAAGCACC TGCCGGAGGA ATAGGGTGAG GGCTGGACAT         37400
GGGCCTGGCC CCCCATCCAT CCTGTCCAGG GACCCCGGCT CACCTGCCAC
TCCTTGCACT GAGGGTCCTG GGCCAGGTTG AACAACGTCA GCGCGTTAAA         37500
AAGCTGCCAG AACTAAGCAG. GGAGGAGGCA TTTTACTCAG AAGACCCAGT
CCCTGCTGGA GCCCGAAACC GGCGCAGAGG ACGGACAGAC AGACAGGCAG         37600
GGACACAGAG GCACGACTTA CGTGTCCAAA GAAAAGAAAA GGCAGCAGGA
AGGTGAGGCC CCGCCACATC CAGGACTGGA AGCCCTCTGC GGGGAGGAAG         37700
GTCAGGGCAC AGGACGGCCA GGGACAGAGG GGACGCTACC ACTCAGACCC
GGGATGGGAG GAAAGGGGAA TGTCCAGATG CCAGGGCCAC TCCCGACTCA         37800
CCCACAGTGA GGTCCATGGT GTGCCGCTCG CCCAGCGCCC GCAGGCGGTA
GAGGCAGCCG CTCTGGTAGT AGTACTGGAG AAACTGCACG AAGCCTGGGC         37900
CGGGCGGAGG ACAGAGGAGC AGGGCTGAGC CCCGGGAGGG GTTCCCTGCC
CCTGCCCCCC ACCACGGGTG CCTCCAGGGC TCAGCACAGA CATCTGGGAC         38000
ACTCACTCTG GTACATGGAA AAGGAGAGGA ATTGGTTCCG GAATTTCTGG
TACATGAGAC CGTCGGGCCT AAGGTAAAAA GTGGAGGGCA GTGTGGGGAC         38100
CCTTGGGGAT GCCGTGTATC CCCTCCTTGT CCCAGCTCCT GCCCCTGCCG
GGGCCCAGCC ACTCACCACG TCAGCATGAC TCCCGACAGG AAGGTGGACA         38200
CGTAGTGATG GAACACCCAC CAGCCTTTGA TCCTGGAGCA GAGAGCCACC
ATCACCCCCA GCGCCTGTTC CTGCTTCCCG GAGGGTCCTG GCACCCCATT         38300
CCATGCTCCC CTCCCTCAGG GCCCGCCCTG CCCACCGGGA GCCGTTGTTG
ATGAGGATGC TCTCCCGGAT GGTCAGGGTG CAGTAGTACC AGACCAGCAG         38400
GAAGTTGAAG GCAGCATCTG TCACCCTGCG GAGGGAGGGG ACCGAGGGTT
GGTACTGCAG CGACTGGGGA TGGGGTGGGT CCTCGGCCGG GGTGGGACGC         38500
CCACCTGGAG TTGAGCAGGA AGCGGCAAGT GAAGGAGATG AGGATGAGGA
```

```
.TGATGGTGAG GTAGAGCTTG AACTTCTCAT ACTCGTCCTT GTAGGCAAAC        38600
CTGGGGGGCG CAGGCCGGTG AGACGGGTGG GGTGCTGGTG GGGCCCATGT
GTGCCCCCAC CCCGGGAGGG CAGCTGAGCC AAGGGTAGTC GGATGGAGGA        38700
GAAGGCCGGG TTGGGGGGTG GAGGGGAAGG CCAGGTGGGG TGGGGGGTGG
AGGGGAAAAC CAGGTTGGGG GGAAAGCCGG GTGGAGGGGT GGAGGGGAAG        38800
GCCGGGTGGG GGGCTGGAGG GGAAGGCCAG GTGGGGGGCT GGAGGGGAAG
AAGGCCGGGT GGGGGGCTGG AGGGGAAGGC CGGGTGGGGG GGTGGAGGGG        38900
AAGGCGGGGA GGGGGTGTGG AGGGGAAGGC GGGGAGGGGG TGTGGAGGGG
AAGGCCGGGT GGGTGTGTGG AGGGGAAGGC CGGGTGGGTG TGTGGAGGGG        39000
AAGGCCGGGT GGGTGTGTGG AGGGGAAGGC CGGGTGGGGG GGTGGAGGGG
AAGGCCGGGT GGGGTGGGGG GTGGAGGGGA AGGCCGGGTG GGGTGGGGGG        39100
TGGAGGGGAA GGCCGGGTTG GGGGGTGGAG GGGAAGGCCG GGTTCCGGGG
GAAGGCTGGG TGGAGGGGTG GAGGTTGAGG GGGGGAGGGG GGTATGGGGG        39200
TAGGGGTGGA GGGTGGAGGT TGGGGGGTTC AGGCTCGGCT GAGTTACTTA
GCCTGCTTGC TCAGGAGCGT GACGTTTACG TTCCCCA                       39287
```

POLYMORPHISMS IN THE CODING SEQUENCE OF POR

```
GGAGACTCCC ACGTGGACAC CAGCTCCACC GTGTCCGAGG CGGTGGCCGA
AGAAGTATCT CTTTTCAGCA TGACGGACAT GATTCTGTTT TCGCTCATCG      100
TGGGTCTCCT AACCTACTGG TTCCTCTTCA GAAAGAAAAA AGAAGAAGTC
CCCGAGTTCA CCAAAATTCA GACATTGACC TCCTCTGTCA GAGAGAGCAG      200
CTTTGTGGAA AAGATGAAGA AAACGGGGAG GAACATCATC GTGTTCTACG
GCTCCCAGAC GGGGACTGCA GAGGAGTTTG CCAACCGCCT GTCCAAGGAC      300
                 T
GCCCACCGCT ACGGGATGCG AGGCATGTCA GCGGACCCTG AGGAGTATGA
CCTGGCCGAC CTGAGCAGCC TGCCAGAGAT CGACAACGCC CTGGTGGTTT      400
     T
TCTGCATGGC CACCTACGGT GAGGGAGACC CCACCGACAA TGCCCAGGAC
TTCTACGACT GGCTGCAGGA GACAGACGTG GATCTCTCTG GGGTCAAGTT      500
CGCGGTGTTT GGTCTTGGGA ACAAGACCTA CGAGCACTTC AATGCCATGG
GCAAGTACGT GGACAAGCGG CTGGAGCAGC TCGGCGCCCA GCGCATCTTT      600
GAGCTGGGGT TGGGCGACGA CGATGGGAAC TTGGAGGAGG ACTTCATCAC
CTGGCGAGAG CAGTTCTGGC CGGCCGTGTG TGAACACTTT GGGGTGGAAG      700
                 T    T
CCACTGGCGA GGAGTCCAGC ATTCGCCAGT ACGAGCTTGT GGTCCACACC
GACATAGATG CGGCCAAGGT GTACATGGGG GAGATGGGCC GGCTGAAGAG      800
CTACGAGAAC CAGAAGCCCC CCTTTGATGC CAAGAATCCG TTCCTGGCTG
CAGTCACCAC CAACCGGAAG CTGAACCAGG GAACCGAGCG CCACCTCATG      900
CACCTGGAAT TGGACATCTC GGACTCCAAA ATCAGGTATG AATCTGGGGA
CCACGTGGCT GTGTACCCAG CCAACGACTC TGCTCTCGTC AACCAGCTGG     1000
GCAAAATCCT GGGTGCCGAC CTGGACGTCG TCATGTCCCT GAACAACCTG
GATGAGGAGT CCAACAAGAA GCACCCATTC CCGTGCCCTA CGTCCTACCG     1100
CACGGCCCTC ACCTACTACC TGGACATCAC CAACCCGCCG CGTACCAACG
TGCTGTACGA GCTGGCGCAG TACGCCTCGG AGCCCTCGGA GCAGGAGCTG     1200
CTGCGCAAGA TGGCCTCCTC CTCCGGCGAG GGCAAGGAGC TGTACCTGAG
CTGGGTGGTG GAGGCCCGGA GGCACATCCT GGCCATCCTG CAGGACTGCC     1300
CGTCCCTGCG GCCCCCCATC GACCACCTGT GTGAGCTGCT GCCGCGCCTG
CAGGCCCGCT ACTACTCCAT CGCCTCATCC TCCAAGGTCC ACCCCAACTC     1400
TGTGCACATC TGTGCGGTGG TTGTGGAGTA CGAGACCAAG GCCGGCCGCA     
                                              T
TCAACAAGGG CGTGGCCACC AACTGGCTGC GGGCCAAGGA GCCTGTCGGG     1500
                                              C
GAGAACGGCG GCCGTGCGCT GGTGCCCATG TTCGTGCGCA AGTCCCAGTT
ACGCCTGCCC TTCAAGGCCA CCACGCCTGT CATCATGGTG GGCCCCGGCA     1600
CCGGGTGGCA CCCTTTCATA GGCTTCATCC AGGAGCGGGC CTGGCTGCGA
CAGCAGGGCA AGGAGGTGGG GGAGACGCTG CTGTACTACG GCTGCCGCCG     1700
CTCGGATGAG GACTACCTGT ACCGGGAGGA GCTGGCGCAG TTCCACAGGG
  G
ACGGTGCGCT CACCCAGCTC AACGTGGCCT TCTCCCGGGA GCAGTCCCAC     1800
AAGGTCTACG TCCAGCACGT GCTAAAGCAA GACCGAGAGC ACCTGTGGAA
GTTGATCGAA GGCGGTGCCC ACATCTACGT CTGTGGGGAT GCACGGAACA     1900
TGGCCAGGGA TGTGCAGAAC ACCTTCTACG ACATCGTGGC TGAGCTCGGG
GCCATGGAGC ACGCGCAGGC GGTGGACTAC ATCAAGAAAC TGATGACCAA     2000
GGGCCGCTAC TCCCTGGACG TGTGGAGCTA G                         2031
```

FIGURE 2

ISOFORMS OF THE POR PROTEIN

```
GDSHVDTSST VSEAVAEEVS LFSMTDMILF SLIVGLLTYW FLFRKKKEEV
PEFTKIQTLT SSVRESSFVE KMKKTGRNII VFYGSQTGTA EEFANRLSKD      100
                                    I
AHRYGMRGMS ADPEEYDLAD LSSLPEIDNA LVVFCMATYG EGDPTDNAQD
FYDWLQETDV DLSGVKFAVF GLGNKTYEHF NAMGKYVDKR LEQLGAQRIF      200
ELGLGDDDGN LEEDFITWRE QFWPAVCEHF GVEATGEESS IRQYELVVHT
                 L
DIDAAKVYMG EMGRLKSYEN QKPPFDAKNP FLAAVTTNRK LNQGTERHLM      300
HLELDISDSK IRYESGDHVA VYPANDSALV NQLGKILGAD LDVVMSLNNL
DEESNKKHPF PCPTSYRTAL TYYLDITNPP RTNVLYELAQ YASEPSEQEL      400
LRKMASSSGE GKELYLSWVV EARRHILAIL QDCPSLRPPI DHLCELLPRL
QARYYSIASS SKVHPNSVHI CAVVVEYETK AGRINKGVAT NWLRAKEPVG      500
                                                   A
ENGGRALVPM FVRKSQLRLP FKATTPVIMV GPGTGWHPFI GFIQERAWLR
QQGKEVGETL LYYGCRRSDE DYLYREELAQ FHRDGALTQL NVAFSREQSH      600
           W
KVYVQHLLKQ DREHLWKLIE GGAHIYVCGD ARNMARDVQN TFYDIVAELG
AMEHAQAVDY IKKLMTKGRY SLDVWS                               676
```

FIGURE 3

POR.ST25.txt
· SEQUENCE LISTING

<110> Genaissance Pharmaceuticals, Inc.
       Kazemi, Amir
       Kliem, Stefanie E
       ·Lanz, Elizabeth M
       Messer, Chad
       Tanguay, Debra A

<120> HAPLOTYPES OF THE POR GENE

<130> POR_MWH-0969PCT

<140> TBA
<141> 2001-10-01

<150> 60/236,449
<151> 2000-09-29

<160> 134

<170> PatentIn version 3.1

<210> 1
<211> 39287
<212> DNA
<213> ·Homo sapiens

<220>
<221> allele
<222> (22635)..(22635)
<223> PS1: polymorphic base G or A

<220>
<221> allele
<222> (29577)..(29577)
<223> PS2: polymorphic base C or T

<220>
<221> allele
<222> (29754)..(29754)
<223> PS3: polymorphic base C or T

<220>
<221> allele
<222> (30427)..(30427)
<223> PS4: polymorphic base C or T

<220>
<221> allele ·
<222> (31531)..(31531)
<223> PS5: polymorphic base G or A

<220>
<221> allele
<222> (31644)..(31644)
<223> PS6: polymorphic base C or T

<220>
<221> allele
<222> (31648)..(31648)
<223> PS7: polymorphic base C or T

POR.ST25.txt

```
<220>
<221>  allele
<222>  (32446)..(32446)
<223>  PS8:  polymorphic base A or G


<220>
<221>  allele
<222>  (32478)..(32478)
<223>  PS9:  polymorphic base C or T


<220>
<221>  allele
<222>  (33538)..(33538)
<223>  PS10:  polymorphic base C or G


<220>
<221>  allele
<222>  (33551)..(33551)
<223>  PS11:  polymorphic base G or A


<220>
<221>  allele
<222>  (33571)..(33571)
<223>  PS12:  polymorphic base C or T


<220>
<221>  allele
<222>  (34766)..(34766)
<223>  PS13:  polymorphic base C or T


<220>
<221>  allele
<222>  (34796)..(34796)
<223>  PS14:  polymorphic base C or T


<220>
<221>  allele
<222>  (34797)..(34797)
<223>  PS15:  polymorphic base T or C


<220>
<221>  allele
<222>  (34850)..(34850)
<223>  PS16:  polymorphic base C or G


<220>
<221>  allele
<222>  (35325)..(35325)
<223>  PS17:  polymorphic base G or A


<220>
<221>  allele
<222>  (35631)..(35631)
<223>  PS18:  polymorphic base T or C
```

POR.ST25.txt

```
<220>
<221>  allele
<222>  (35632)..(35632)
<223>  PS19:  polymorphic base G or T


<220>
<221>  allele
<222>  (35721)..(35721)
<223>  PS20:  polymorphic base C or T


<220>
<221>  allele
<222>  (35774)..(35774)
<223>  PS21:  polymorphic base T or C


<220>
<221>  allele
<222>  (35992)..(35992)
<223>  PS22:  polymorphic base C or T


<220>
<221>  allele
<222>  (36054)..(36054)
<223>  PS23:  polymorphic base C or G


<220>
<221>  allele
<222>  (36207)..(36207)
<223>  PS24:  polymorphic base C or G


<220>
<221>  allele
<222>  (36330)..(36330)
<223>  PS25:  polymorphic base G or A


<220>
<221>  allele
<222>  (36617)..(36617)
<223>  PS26:  polymorphic base T or C


<400>  1
cagtggcaca atgtcagctc actgcaagct ctgccttccg ggttcacacc attctcccac    60
ctcacctccc tcctcctgtt tctgggacta caggtgcctg ccaccacgcc tggctaattt   120
cttgtatttt ttttttttaa gtagagacgg ggtttcactg tgttagccag gatggtctcg   180
atctcctgac ctcgtgatcc acccgccttg gcctcccaaa gtgctgggat tacaggcatg   240
ggccaccgtg cccggacgac ctgacttact ttttaaagca tcccccctgtc tgctgtgtgg   300
ggaatggccc gtccatcaga attctcttgc ggtgattacg taggtgcaga tgtaccgcta   360
acctcactcc cagcgtcgtc gatttattaa tagaaactgc catgagttat gcaatcagct   420
tgagtctgcg ttgagaagga cttcacatca tgggcattag atctcttctg gaaaatgagg   480
ggaccggtca cttggggagt tagaagggag gagggtgagg gacagggagc gcgagtttct   540
tggctgctgc cttgtgatgg ataacagcta gacagctagg ggagctcctg tgtcaccatg   600
gcactcctga ctggttccag tggggccctg actgtctgct gcctcatgac tctgtgggac   660
aagtccttcc agaatctgtc agagcagatc ctaaatgcct ctctgtctgg ttctatgaaa   720
atgtgacagg tggctcataa gcttgacaga gatcctgata tgactagcat taagtaaata   780
ttaaatgtgt ttataagatt aaatataaat agttttctct ggggacaaca gaattcactt   840
tgtcctttaa tgtcatatca aaaaagctgg ccaggtgcag tggctcacgc ctgtcatccc   900
agcactttgg gaggccgagg caggatcacg aggtcaggag tttgagacca gcctgaccaa   960
catggtgaaa ctccatctct actaaaaata cagaaaatta gccgggtgtg gtggcacatg  1020
cctgtaatcc cagctgctca ggagactgag gcaggagaat tgcttgaatc cgggaggcag  1080
aggttgcagt gagccgagat cgcaccactg cactccagcc tgggtgacag agcgaggctc  1140
```

Page 3

POR.ST25.txt

```
tgtcccccgg cccccccccc ccccgaaaaa aaaaaagtca tatcaaaaaa gccaagttga   1200
ggccaggcac agtggctcac gcctgtaatc ccagcacttt gggaagccga agcaggcaga   1260
tcacctgagg tcaggagttc gagaccagcc tggctgacat ggtgaaaccc cgtctctact   1320
aaaaatacaa aattagccag gtgtcgtggc gggcacccgt agtcccaggt accctggagg   1380
ctgaggcaag agaatcgctt gaacccagga ggcggaggtt gcagtgagtc gagatggtgc   1440
cattgcactc tagccagggt gacagagaga gactctatct caaaaaaata aaataaaaaa   1500
gaaagaaaaa agctaaattg agatgaactc aagccattgg aaaagcagct cagattagtg   1560
agaaagcaaa gacccaagct tgcgggcctg ttctgaatgg cctgttccca gcaggttgtg   1620
tagcagtaga ccgatgccca caaccctcgg atgccctggc ttcccagacc agccccatgg   1680
tttcctggtg acagtgacag ccatcagcag tcagctagaa catagtctct cattgcctaa   1740
ttgctgacac atgactaagg ctttcctgat tgagtctggg tgctctggtt gtggtgaaaa   1800
acccagtcgg tcatatccat gattgcaaaa tggaatagtg tgaatttcag ggctgagccc   1860
agcctccttg gtaagcagca gaagctgctg tttcagggac tctcacttgg acctttgttg   1920
actcatgtca cgtctgccta ccaggtgtcg ttggtggaag gaccagttga gcatcttatg   1980
agggaactag gctggtgaca tgggggttcc cccacctccc tcctggacgg ggcggctggc   2040
cggggcgggg gctgaccccc ccacctccct cccggacggg gcggctggcg gggcagaggg   2100
gctcctcact tcccagtagg ggcggccggg cagaggtgcc cctcacctcc cggatggggt   2160
ggctggccgg gcggggggct gaccccccca cctccctccc ggacggggcg gctggccggg   2220
cgggggggctg acccccccac ctccctcccg gacggggcgg ctggccgggc gggggggctga   2280
cccccccacc tccctcccgg acggggcggc tggccgggcg gggggctgac cccccacct   2340
ccctcccgga cggggcggct ggccgggcgg ggggctgacc ccccacctc cctccccggat   2400
ggggcggctg gccgggcaaa ggggctcctc acttcccagt aggggcggct gggcagaggc   2460
gcccctcacc tcccggacgg ggcggctggc cgggcggggg gctgacccccc ccacctccct   2520
cccggacgga gcggctggcc gggcggggggc tgaccccccac ctccctcccg gacggggtgg   2580
ctgctgggcg gagacgctcc tcacttccca gatggggtgg ctgctggacg gaggggctcc   2640
tcacttctca gacggggcgg ttgccaggca gagggtctcc tcacttctca gacggggcgg   2700
ccgggcagag acgctcctca cctcccagat ggggtcgcgg ccgggcagag gcgctcctca   2760
catcccagac ggggcggcgg ggcagaggcg ctcccccacat ctcagacgat gggcggccag   2820
gcagagacgc tcctcacttc ctagatggga tggcggtgga gaagaggcgc tcctcacttc   2880
ctagatggga tggcggccgg gcagagacgc tcctcacttt ccagactggg cagccaggca   2940
gaggggctcc taacatccca gacgatgggc agccaggcag agacgctcct cacttcccag   3000
acggggtggc ggccgggcag aggctgcaat ctcggctctt tgggaggcca aggcaggcgg   3060
ctgggaggtg gttgtagcga gccgagatca agccactgca ctccagcctg ggcaccattg   3120
agcactgagg gaacgagact ccgtctgcaa tcccggcacc tcgggaggcc gaggctggag   3180
gatcactcgc ggttaggagc tggagaccag cccggccaac acagcgaaac cccgtttcca   3240
ccaaaaaaat acgagaacca gtcaggcgtg gcggcgcgcg cctgcaatcg caggcactcg   3300
gcaggctgag gcaggagaat caggcaggga ggttgcagtg agccgagatg gcagcagtac   3360
agtccagctt cggctgggca tcagagggag accgtggaaa gagggaggg gaggggagg    3420
gggagggacc agactttgtc tcttaaaaaa aaaaaaaaag actcccaggc ataggaaagc   3480
aggcattcac cataaatctc attgttaaca ggaactacct tcatagccca aggttttagg   3540
aatacagaga cactcttgcc agtcaggata gtccaaggct cgtgaatatc aggagccact   3600
caagggccag gcctggagac ttggaacatg cagggtttag ggcaaccccc gccagccaca   3660
cagctgggaa aagggtttgc tgcgaaatgg tccactcttg cgtgcacttc ctcccgcctg   3720
ccccccgtgc ggtgtccttg tgggattaga ggatgttcca gcacactgag agtcagtggc   3780
tgtcattttc tctgtcttcc agggacccag tcccagctgc tgccttagct tgcaccccag   3840
gaccaccctg ggctggtgct gacaccactg ggggctgtct gttatgtcag ccccagtcca   3900
agactgtggc tgtcttggca ggaacctggc tgagtgagcc ccttctccta ccccgtgcag   3960
tgaccatttc ctgcagcccc aggggcaagg cccagcattt aggtgggcac cccagccagc   4020
ctcaggggca ccctgctacc ctctgctgac atctgctgtt tctgtctcct aacagtttca   4080
tgatcaacat gggagactcc cacgtggaca ccagctccac cgtgtccgag gcggtggccg   4140
aagaagtatc tctttttcagc atgacggaca tgattctgtt ttcgctcatc gtgggtctcc   4200
taacctactg gttcctcttc agaaagaaaa aagaagaagt ccccgagttc accaaaattc   4260
agacattgta agtgccgcct ctcagcctcc tctctctgtc cctcttctgt caccactcca   4320
agcagtgtcc tgcatgcggg cctcaggctg aaagaagcaa ggctccttgt agcattttgg   4380
ggtgactctt gggtttttgcc ttcttgctag attgtgcttg agctctccct taatcatctc   4440
ttatgtgggg tttataaggc ccttttctat ctgctgttta tatgaccttc acaggaagca   4500
ctgcgaaaat gacaccaagtt tacagattaa gaaccggggg ctcaagaatg ttaacggggcc   4560
tgccagggtc attagccagc aatagcagcg ctagagctgg aacccttttt tccatttatt   4620
tatttatttta tttatttgga gacatggtct cactgtattg cccaggctgg agtgcaatgg   4680
cggtggcacg atcaaggttc actgtaatct ccacctccca ggctcaggca gtcttcccac   4740
ctcagcctcc cgagtagctg ggactaaagg tgtgcaccac tatgcccagc taatttttttt   4800
ttttttttttt ttgaagcaga atcttgttct ttcaaccagt ctggagtaca gcggtgcgat   4860
ctcggctcac tgcaacctct gcctcccagg ctcaagcaat tcttctgcct cagcctcctg   4920
agtagcttgg attacaggtg cccgccacca cgcctggcta ttttttgtat ttttagtaga   4980
gacagggtct caccatgttg gccaggctgg tctcaaactc ctgacctcaa gtgatccgcc   5040
cacctcggcc tcccaaagtg ctgggattac aggtgtgagc cgttgcgccc agtcaatttt   5100
tgtatttttt tatagagaca gggatttgcc atgtcccggg ctggtctcga acacctggac   5160
tcaagcaatc cgcacacctt ggtctcccag agtgtcggga ttacaggcgt gagccactgt   5220
```

Page 4

```
                              POR.ST25.txt
acccagccta gaattggcat tttaaagtct ttttaaggct gtagccacct ggaaggctat     5280
gtgctaatga gcacaaaatc tttggacagg ccttattttg ttgctctttt aaacttttca     5340
tgaagcttcc ttttaagaat aagtgaaaaa gatggattag ttaaaagaaa aacgttaagg     5400
aaataatagt atagatggta tgtagacaga aaaaaatgat accgaaggtc tgcctaagtg     5460
actgcatccc tgtctctcca agattgattt ccataagcca cacggaaaag tcattctcca     5520
gatttaggg tcctgcggcc agaattactt cctgcggtta cgcttggctg tgtttaacac       5580
cagtctcaga ctccaggggc ccgcagccag gtttgagtgg ggacctgcca ggtcaggtct     5640
gagccgaggg aactccaggg gcacgggcct gctgccttca gggactccct ggatttggca     5700
aggggcaggt ctcaggggca ggattgggag ccattctcag accagaaccc ctttcattct     5760
gtctgtcttc aggggaccca gtcccagctg cagccttagc ttgcacccca gggccaccct     5820
gggctggtgc tgacaccagt gggggccatc tgtcatgtca gttccagtcc aaggctgtgg     5880
ctgtcttggc aggaacaagc cacacagaaa agtcattctc caaggctggg catggtgaag     5940
catgccgctc acatcaaatc ccgattccct attccaatct gggtatcagc cttccccgcc     6000
agaacactta cagccttcac accgtaaagc agatattcac acaacacagt tcttatacca     6060
tggacaggga attagtttta cattttaaa aactgggttc agggccaggc acagcggctc       6120
acgtctgtaa ttccaccact ttgggaggcc gaggcgggag aaatcacttg aggtcaggag     6180
ttcgagacca gcctggccaa catggcgaaa ccccgtctct accgaaaata caaaaactag     6240
ctgggcgtgg tggtgcatgc ctgtaatttc agctactagg gaagctgagg cagaagaatc     6300
gcttgaaccc aggaggcaga ggttgcagtg agccaagaat gcgccactgc attccagcct     6360
gggagacaga gtccgtctca aaaacaaaac aaaacaaaac aaaactgggt tcaatcttat     6420
ctccatcatt gactcactga gcaaacattt cctgaggccc tacctacagt gtcccagaca     6480
ctatgctaag gtgtgtggct gcagcagtga agaaaataga tgaatgctgt cttcatcaag     6540
cttacgctct agtggtgagt gaaccctggg cgagacctta acctcattct gcctcagttt     6600
cctcacctgc taaatggagg cagtattccc aatcttgtgg ggcttttgtc ttaaatgagg     6660
aagagtctat aaactcacct agcactgccc ctgagggata ggagctgtgt agtgtttgcg     6720
ttgatcgtat ttcgtgccag gcactatgac gggcacgatc aggaccacag agccagcatt     6780
tatgttaagc tggaaaatga gggtgttgca acatagtttt tttttttttg agacagactt     6840
ttgctcttgt cactcaggct gggagtgcagt ggtgcaatct cggctcactg caacctccgc     6900
ctcccgggtt caggcgattc tcgtgcctca gtctcccgag tgcctgggac tacaggcaca     6960
tgctaccacg cctggctaat tttgtatttt ttagtagaga cgggtttca ccatgttggc       7020
caggctggtc tcaaactccc gacctcaggt gagctgcctg cctctgcctc ccatagtgct     7080
gggattatag gcgtgagcca ccacacccgg ccacagcgta ctttttagaa atgctttgtg     7140
aactgcttta acctacattg tctaatcagg gcctccttgc aacagcgccg tgcgttaggt     7200
agcgtgggtt gcctcatttc tgctctcaga gaggttatgg gatgtgccca gggcacaggc     7260
tggagagaca ggctccgcca gctctaagag agaaggaatt ggtgtgggga tcagaccaag     7320
tcctgcagga ggggtgttgg gcaggttctt aaaggtagag ccacacccty aagctgaaag     7380
gactttagag cagagggccc gatataggaa gggcagatca gctggaatcc agcccatcct     7440
gtccagggct ggtacacggg gcattgctgg tgtggaatag tactttgat aaaaaaagcgt      7500
agtagtactc aagataaaga gaccagggtc caggacccag ctcctgaact ggcccttgtt     7560
ccctcatctg taaaatggaa gttccgcccc accgccgtga gtgaggacgt gggttggccc     7620
tgggagcctg cggagcttgc caccaaggga ggcagcattc ctctcatagc tcatctgttc     7680
tcgctgtcac ctccctggcc ttacaagagg aagttgacat aaaacctgta gactgtggga     7740
gctgtaggt cccaaagcac ttgtattttt ttgttgattc tcactgttcc tctttgccag       7800
gatgctgaaa ttctgttccc tccacctaat catgatcttc taaccttatg tgtgtggaga     7860
tatgcagttt ccagagtctt gtctcttccc tcattaaatc acactggatg cctgctgcag     7920
ccagggcttt tacctgtacg ctgtacaaga caaatatttc tagtaccatg gctacattct     7980
ggaccctgaa gttgggtgct caaggttgag tactgactct cccacattag ccatctgaca     8040
tcagccaggt gacttgatct ttccaagtct gtgtccccaa ctgtaagctg gaggtagtag     8100
tagtacttac ctcataaggc tgtgaaaact gagttaatcc atcatagagc tctttgctca     8160
gttcctggtg catagtaggc tctcaggtgt tatggattgt tactatttat ttatttattt     8220
tgaggcaggg tctcactctg tcgttcaggc tggaatacgc tggcacgatc tcggctcact     8280
gcaacctccg cctcccaggc tcaagagatc ctcccacttt ggcctcccaa gtagctggga     8340
ctacaggcgt gcaccaccac ccttggctaa tttttatatt ttttgtacaa atggacatct     8400
aacaatgttg cccaggctgg tctgaaactc ctgggctcaa gcagtctgcc tgcctcggcc     8460
tcccaaagtg ctgggattac aggcgtgagc caccatgcct ggagtgtgtt attattttta     8520
tccttttat taataccatc tccagtctta caagagccct tccacgaaga catgagcagc       8580
ccattttgcg gatgaagaaa ccgaggtgcc agggattgca gagtttgccc aaggccacct     8640
aacttctaaa tagcagcaga gttggctttt gaatgaagtc tgcaagctcc acaagccatg     8700
ttttctctac tttgccatgc tgctcctaat ttagtgttaa taggcctcct acgatagtgt     8760
tagtctccta agatagtgtt aataggcctg gcatggtggc tcacgcctgt aatcccagtg     8820
cttagggagg ctgaggcaca aagattgctt gagccaggag ttcaagacca gcctgagcaa     8880
catagggaga cccccaccc ccgccatct ctacaataaa caaaaaagag tggtgttaat        8940
ttgttaattt accgacaagg agagtattcc ttatctgaag tgcttgccag aagtgtttca     9000
gatcttggac tttggaatgt ttgcattata ctcaccagat gagcatttct tatctgaaaa     9060
tctgaaatgc tccaatgatt cttccctttg aggatcatgt tgacgctcaa aaagtttcag     9120
atttcagatt ttagattagg gacattcagc ctgtagatgt tattttacag agcaggaagt     9180
tgggtgtagt ggttcacact tgtgatccca gctctttggg aggccaaggc aggaggattg     9240
cttaagccca ggagttggag accagcctgg gcaacgtagc aaggccctgt ctctataaaa     9300
                                Page 5
```

POR.ST25.txt

```
aataggagaa aactgaggct cagaaaggtc acctagctga tacaaccagg cctcaagcct    9360
agatctccta ttctaaatcc cacatcagtg tccctggcta ccggtatttc agtgagcaca    9420
ggtgatttct gttatagcca cagagcagta gggagtaata gggcttagtg tgaggaaggt    9480
atgtggccca gggctatatc aggaactctc accatgattt ctggtggctt ctagaattgg    9540
catttattta ttttgtttta gagacacggt ctcgccctgt cacccaggct ggagtgctat    9600
ggcgcgatat cagttcactg caacctctgc ctctggggct caagcgattc tcttgcttca    9660
gcctcccaag ttgctgggat tacaggcacc caccatgatg cccggctaat ttttgtattt    9720
ttagtagaga tggagtttca ccacgttggc caggctggtc ttaaactcct gaccccaagt    9780
gatccatccg cctgggcgtc ccaaaatact gggattacag gtgtgagcca ctgtacgggg    9840
ccaaacacat ttttaaaaca aaaaacccag tgggattttt gttggaattg gattgaatct    9900
gtatattaat atgggaagag ttgacatcct acaatattga gctttctatt ccatgaacat    9960
tgtttttatcc ctcccgggga ggtcttttgg actgtttgta atagcaatta atgacttgtt   10020
agagaggagt ggtaatgagc aacaagttct tttcataat aatagtagca gtaatgataa    10080
taaatggctc ttcctgccag gtgtggtggc tcacgcctgg aatccacac ttcaggaggc     10140
cgaggcaggt ggattgcttt gagcccagga gttcaagatc agcctggcaa catgacgaaa    10200
ccctgtctct acggaaaaaa tacaaaaatt agccaggcac ggtagcatac gcttgtagtc    10260
ccagctactc aggaggctga ggtgagagga tcacctgagc ctgggatgcc aaggctgccg    10320
agcgccgagc tcacaccgct gcactccagc ctgggcgaca gagcaagacc ccctctcaaa    10380
ataataataa taataataat aataatatat aataatgact ctttcacgtc cactcccact    10440
cactcaggtt gcacctctga acagagcatt tcgcactggg gcagctcatg ccctaaggga    10500
gccccaggga atgctcagcc tcttgagtag ctggggtgac gggcacatgc caccatgcct    10560
ggcttattta tttttttta aagacatggt ctcgctctgt tgcccaggct gggtcttaac    10620
ctcctggcct caagcagtct ccctgcctca gccttgcaaa gttctgagat cactcactgt    10680
gtaggatcca aagtcccaca ggatccaggg ctcctgtggg ctctctctcc tgtgggcctg    10740
gcagcttgc tcacccttat ggaacaacac cgcatggcgt gccctcttgg tgatggaatt     10800
tgtattttg cctcccatgg tgcagagagc gtcccatttc catctgggtc cctaccttag     10860
tgcggggccg cctgtcggag ggaagcttct cagagaatgg ccgttgaatt aaccaaggct    10920
aaatctgtat gtgtggctgc ctctagggaa acctgtggcc tccaggctgg gtttggctta    10980
cacagtattt ttttaaaaaa tattttaatt agaacattaa aaagtgtggc agtatcaggc    11040
ccagcgtggt ggctgacacc tgtaatccca gcactctggg aggccaaagc aggtggatca    11100
catgaggcca ggagttcaag accagcctgg ccaacatggc aaaacccgt ctctacaaaa     11160
agtacaaaaa ttagctgggc atggtggtgc gtgcctgtaa tcccagctac tcgggaggct    11220
gaggcaggag aattgcttga atccaggagg tggaagttgc agtgagctga gatcacgcta    11280
ctgcattcca gcctgggcga tgaagtgaga ccaaaaaaaa aaaggcagta tcaaataaaa    11340
acttagactt acagcttctt taaaaaaaaa aaaaaaata gaggatctgg tcatactgga    11400
cccacagtga tgcccaccag ctgtgggagca gctgagcagc ggccccattc tgcacgaggc    11460
ctgtgcgctg cattcactct gaccctccag cccgctcagg ctgctgtgct catgttgcct    11520
ctactggccc tgagaagctc tgcagtgccc cctgcacctg tgctttgctc ctggaatgct    11580
agagtgcttt aaagaaaaat aaagcttctt atggaacatt tcaaacacca gagtgtaggg    11640
cacaggagag tggaccccat acgcctccag cccccagttt tcataactac tcactcctcg    11700
ccggtcttat ttctttccgc tgcccccac cgccgcacca gtcctggatt attttaatgc     11760
aatccccaga tgtcatattt catcaggaaa tgtttcagtc ctctaaaatt caaggactgg    11820
ccaggctgg cggctcatgc ctgtaatccc agcacttcgg gaggctgagg caggagaatc     11880
acttgaagcc aagtgttcga gatcagcctg ggtagcatag caagacccg tctctacaaa     11940
aaacacaaac attagtcaag tgtggtggca catgcctgtg gtcccagtta ctcagaaggc    12000
tgaggtggga ggatcacttg agcccaggag gttgaggctg cagtgagcca tggttgtgcc    12060
actgcactcc aacctgggtg acagaacgag accctgtctc aaaaaaaaaa aaaagtaaac    12120
ttcaagtatt ctcctttctg aacataacac aacgccatct tcacaccttg gaagtcaaca    12180
gtaattcttc agcatcgtta gatacccaga caacatttat tcatagttcc ctgattgact    12240
gatgtttgat tttgtttaac agtttttgtt tgtttaattc aggatctaaa caagacccat    12300
acactgcagt tgattaataa atctttcgtt tgggagaggt caggtatatt caggtataat    12360
ttgcatacag taaaatgccc ccctcctccc ttttttttga gacagggtct tgctctgtca    12420
cccaggctgg agtgcagggg tgtgatcagg gctcactaca gcctctacct cccaggctta    12480
aggggatcgt cccatctcag cccctgagt agctggggct acgggcatgt gccaccatac      12540
tcggctaatt tttatttta atttagtag agacaaagtc tcactatgtt gcccaggctg      12600
gtcttgttcg aacgtctgag ttcaagtgat cctcttgcct cagcctccca aagtgctagg    12660
attaggatta caggcatgag ccactgcact gggccaacat tttcctcttt aagtgcacca    12720
gttcatgagt gtacagat gtgtgatt ctgtaatcag tatcataatc gaaatctaga        12780
acatttcatc actccgaaac agtcctgtcc tacctctttg caatcatcct cctccatcgg    12840
gtaatcagcc tctggcaact gccaatctta tttctgtccc tgtagtttaa cattttccaa    12900
agcgtcacat aaatggaatg atagcaccgt gggctgctgt gccagcggtt tttccttagc    12960
gacacgtcag ctccctttcc attcctaagc ttcccttcca tcccgtttct tccttgaaac    13020
atcatttgta tcacagtttc ccacagtctg gattttcctg attacgtccc cctggagatt    13080
gcttttgaaa agccacaaaa gatctcattt tcctgagtga ccttcctttt ggggcctgtt    13140
cttgacagaa agcccgggag gccaggggtc agggctgcca gcatcacgtg tctgttttt     13200
atctcctgag tgccaacaga tgggaagtgg atctggggt cttgccacat ccagttcagg     13260
gtcactttct ctgtagtgag gtccttggct ccatcgagca tgaaaagaga aatacaaatg    13320
agactgacaa agaggaaagc attaggggaa gcctgtggcc acgagagtag acggtgcagg    13380
```

POR.ST25.txt

```
gggcccactg cctcccgctg ctgtggggaac ggcgaggaca tgggccaggg tgtttctgcc   13440
acttctgcaa gcacagctct gggtttttctg ctgccatctg acctccaaag ggacctcgcc   13500
tcctctcctt cctttcctgg tccctgatga gatggggggag gtttctgtca ctctgacaag   13560
tgcttgaaac tccagacatg ggctgtcggg aactgtcggg tactgccgtg cagagcgcat   13620
gttgtctgac ctcctgactt cactgggccc tgtcgggtgc aggagaggga gctgtggatg   13680
gcacagtggg acatgtgctc tactggcctg ggggtggtgg gaatgcagag gaggggtatc   13740
cccattgggg gtgtgtcagg gatgagatgc ccccacagga gcctaggagg ccaagtgtag   13800
aggctgggag gtgaggacac gcggggctgt tccaagggag ccacatgctg gtagagccca   13860
ggcacaggtc accaagggag tcaggggtcg tgggggggtca gaggtcataa aaggagttgg   13920
ggtcaccaag gaagtcaggt catggaggga gctgacacag gagagcatct tagagggcct   13980
tgggaccttc caaggggtat ggattttttt ccccagcttt attgaggtat aaatgacaaa   14040
aattatatat atttatcaga gcatctaagt ttttttcttt ttctttcttt ctttttttt   14100
tttgagatgg agtcttactc tgtggcccag gctggagttc agtggcacaa tcttggctca   14160
ctgcaacctc tgcctcctgg gtccaagtga ttctcctgcc tcagcttcct gagtagctgg   14220
gactacaggt gcctgccacc acgcccagct aatttttgta ttttgagtag agatgggggtt   14280
tcaccatgtt gaccgggctg gtcttgaact cctgacctca agtgatccac ccacctcagc   14340
ctcccaaagt gttgggatta caggcgtgag ccaccatgcc cagcctaagc attttttcttt   14400
ttttctttt gagactaagt ctcgctcctg tcccccaggc tgaagtgcaa tggcgcaatc   14460
tcggctcact gccacctccg cctcccggct tcaggcgatt ctcctgcttc agcctcccga   14520
gtagctggga ttacaggcgc ctgccaccac acccggctaa ttttttgtatt tttagtagag   14580
acggggtttc accatattgg ccaggctgat ctcgaactgc tgacctcagg tgatccgcct   14640
gcctcagcct tccaaagtgt tgggattaca ggtgtgagcc accgccaccag gccagcattt   14700
ttctttcttt gtaaattttt ttttcccag ctttaaaatg acataattga caaataaaaa   14760
ttttatacat atatttaaaa ttatgtaagt tcaaagtata cagtgtgatg ttctgatata   14820
cgtatcttag ctgagcatgg tggcttacac ctgtaatacc agctacttgg gaggctgagg   14880
cgagaggatc acttgaagcc aggaggtcaa ggccagcttg ggcaacatag caagacatcg   14940
actctacaaa aacttaaaaa taaattcaaa aattagctgg gcatggtggt gtggacctgt   15000
catcgcagca ctttgggagg ctgaggcagg agttggaggc tgcagtgagc tgtggtcacg   15060
ccactgcact ccagcctggg caacggagcg aacccgtct ctagaaaaac taaaaaatca   15120
atttaaaaa atgacatatg tggatgacgc taacagggag ctggggaagc ctgtgtagga   15180
tggcggtgtc tggagagcgg tggccccatg aagcgccatt tggagctgga gtggagggag   15240
ctgacttgga agcaggaaga tgggccagga ggctgctggc ctcagcctag cagaaattga   15300
ccagggccgc tcggcacaga tggagagccg agtgcaaatg caggtgtcag tggatcagag   15360
aaagggctcc aagccagtgc ttttgaaaac ataattctgg gaccctcatc acccaccaga   15420
ggtgatgttc ttatctccat tttacaaatg gcagtgctga tcccaggaga gggtgagtga   15480
cctggtcaag gctgacgtta gcacatgaga gctggattc cagcctgggg cttcctgcac   15540
cacgcttttg ctccctcgga cccctagctc ttgtcatggc cacccagcc caccectcct   15600
ttccatcccc tttgattgca cacaatagca gagccacttt gtgccagatc agatgatctt   15660
actgtctgag ggtatggcct tttctttctg tcaccgatgg gctgaataac cctgagtttc   15720
tcttctctgt ctccggaaaa ttagccctag tcctggccac cccctggctc agggtcactg   15780
ctgaggtcgt cagaagagca tcaaggcttt tgcgtgtgtg aacagcaca gcaggctttc   15840
tgagcttcac agacttgctc acggagtcct gatgtgaaca cgtgctggag gacatggtcg   15900
gggggaatga gggagcggca gcccaccaga gccttagagc atcctgagga cacggcccca   15960
ggatccttgg gaagggctgt gagatgtgct gcctgaacgt gcccgggcg cactgcagac   16020
aggtaaagc cggggggaggt gcaccttgct tggggagcag acaggaaccc cgggtgctgg   16080
gccactgggc cctggagctt tcacaggcct ggagttgaat taccgcttca cccaccctga   16140
gcctcactgc cttgtcaagg tggatagaaa cccttgcgca ctgcacaggg tggtgaggat   16200
taaacaaat aaaatagaaa gcttcttgaa tgctagccag catgcagtag gcactcagta   16260
agaagtcatt ccttcgcct gttttgtctc tgacctgtgg ccactcacac ctgctttcac   16320
ctcccctgc ctgtagccag ctgtcctccag cagctgccct ccccacgtgt tcatcagagt   16380
ccctgtccgc agagacagct gctggcttct gtctgagctc cagcaaacca ctttctgttt   16440
tcaccaaaga ctggcgtttc caaccattag ccccatgcca caaggctctg aaggtcaaga   16500
aatctgggtc cgcctagcag agccactct ccgtcctgcc agagcaggct gcccccagct   16560
ctgctgccat cgtgcctctg tgcctcagcc gccacccctt ctctcgccgt gggcacccat   16620
gaccctggag cctggagaac ttggcagtag acgctaggtt ctgaaaggga gctggtccaa   16680
agcagcttcc ccatgtgttg ctaagggcag gggacagtct caggagaaat cgagtccgct   16740
gcagaaaatc caaccggatt tgcagtttcc agaaaatgttc ctcagcgtcg cctgtgacgg   16800
gaacaaaatg aaccgcacat ctcggttcat ctcggtccaa gtgactcagt gcttttcccg   16860
ccctccctca tcttccccac gcagtggact acagctgact tctgatcctg aacccagaaa   16920
gactcctgct cccaatttag gcgtggagag ccttaattgg aacctttcat gctccgttct   16980
ggagccacgt ggtctgtgtg ctccaggcct gggacagctg ccacctctgt acggctggag   17040
ccttgggctt tctggttttg tttgtttttt cttgagacag ggtctcactc tgtcaaccag   17100
gttggagtgc agtggcacaa tcttggctca ctgcaacctt cacctcagcc ttcaggctca   17160
agtgatcctc ctacctcagc ctcccaagta gctgggatca cagacgcaca cctctacacc   17220
cggctaattt ttgtattttt ggtagagacg ggctgtttcg ccatgttgcc caggctggcc   17280
tcaaactcct gagctatggc gtccacccac ctcggcctcc caaagtgctg tactggtgtg   17340
agccactgca ctccgccgag ccttgggctt tcggatgggc tgtttgcact cgggaacctc   17400
agaagggtgc tccccaagac agaacatcac cactgcaaac ggagcccacc ttgccaggca   17460
```

POR.ST25.txt

```
tttccccaac tgccctgatg ggaggggaag gtgtgtcttt catggcacac aggatgcttc 17520
ccgctgacat ggacccagtg tcttaagctc tgtttgcccg caattctttt gtgtggggga 17580
agacagggac cagacttaat ttaatgtgat tgtgtggcgc cctggcagtg ggatcagtgt 17640
ggcttccttt atggagtgca gccctgccat agcaacctgc aggctgtctt ctagagccta 17700
gaagagcttt gttccgcaga catgacttcc tttttttttt taatgggaaa cggggagggc 17760
gtggggagaa gggctctgat ataaacagtt accaaggaga cccgagtcgg ggttggaaag 17820
gttttctcct tccctcggtc catacagagg gaggaggggt gtgtgcccgg tgcagggact 17880
catgggaggc ttcactctgg aatcctgagt cctgcctgct ggggctccag atgttgcaga 17940
gttcttgggc tcatgggggc agacgcaggc aggagctcag ggaaattccg agaggacttt 18000
ggggtgtcca ggagcacttt gaggccacgg gcccaggcaa cttcctcacc agtacggaac 18060
aggggatgca cagggccgaa ctttccatct gctggcccct gctggtgtcc aggagagagc 18120
tgtgaccctc ctttttcctt aagtcctgga gctgagagtg agaagccatg gggtgcacgt 18180
actcgtctcc acctgaggag cccgctctga gaaggtcagt tcctgacact gtccctggcg 18240
gcagactgcc ctcatgcccc ttggccaggg gacgcctgca ccctcactcg agtgtgggag 18300
acttcctcct ggcggggacc tgctggaggc ctgaagatgc ggctgcctgg agttaacagc 18360
aaggtcagcc attcccaggg ttgctgcccc tgtctctggc gagggactca gccaggccta 18420
ggcactggtg gcccctctgg tccaggactt gcacgctgca tttccctgct tccaaaggaa 18480
ggaatttgga ctggagacca ggaaggcctc atgcaccctc ttctctgcct ataaataatc 18540
aggtccccca tgcaaatcca ggtctgaggt cctccgatgg aatggagccc attcagagca 18600
gtccagtgtg ggccctctcg ggctaagccc gtgaactttg cttcttggag cccccagcag 18660
ggccttgggt tgggaaggga gtttcctggc accctccccg gcaggatggg tttaaccagc 18720
aggaacttga aggaattgtt ctgtgttgcg ccggtgacct agaggcacag gtccgcctcc 18780
agctccttgg gttccctttt cccccctgga ccactcacac ttcctgggtg aaccctagcc 18840
ttggggcaga gcaggctggc acaccttgtc cactctgacg ttccagccca gaggctcccc 18900
tgtccactcc cagggccagc cagcctggca ttctgctttg tccaactgcg tcaggaggtc 18960
acatggaagc cctcctcccc cgccacttgc cgggcttctg ggatgggagc cgagcttgga 19020
agttgccatc ccttggcacc ccgcttcccg tggggctgtc tctccagggg gccgtttct 19080
attcctgaga agacagccca gctcgtccaa gaggcctttc ccggtttctt cagccataag 19140
cgtctgctcc cttgaagcc ctgtggcgtg aacctgtcct gtcccatccg tccctgtgg 19200
gtgaactgtg ccatccatct atgtggtctt agcacatgcc accttgtgcc atggatgtct 19260
ttgaaatcac aggctgaccg ccttccaggg aatgcccttt tttctactgg gtgttaacca 19320
agttccccag tcaggcactg gggaggctgc aataaccaga tgcagctctt taataatagt 19380
gtgggcacga gaggggcggg gcctgtgcag tcccgtgcct gtctagtgtc agagccagga 19440
tttaaagcca ggtccgccct gccacgtggc tcttcccatc ttgctgggat cctcctcgga 19500
ggctgtcttc ccctggctgg aggattgggc ctgtgaagtc acaggttgtc tcgttctagt 19560
ctcatgccct cggtgcctgg cgcagtgcct cccgtgcccc cacctggctg gcctgtgtta 19620
cccgccaggc actcacgggg cactggagtt tgctcttgtc cttttcttca gcaaatcttc 19680
agatccactt tgttggagat ggaacttccc tggaagcatc tgtcctttga gtagagatag 19740
gggttggctt ttcagctgat gccaggtcag acctcagagg tggctaggga gtgagaagtg 19800
ccctcaacag cctcttagtc cgcccctgcc agcctcgtgc tggacagaga actcagggct 19860
ggccacccta ggaggagaca cggcgtcctg tgccgtgtgc taggcgtcgc gctcatggca 19920
gggatacaca gacagccccc aaggttattc aacccaggct tccagttaag ccacaggaat 19980
gtgcagagta aagtgccttg gccgaccagt gtgtgtgcca cagcgaagac ggcgacagta 20040
gcgttgctgc gtgagggca gggacggcct ctgggagagg aaagttaagt tctgtggtaa 20100
aggatgagta gaaggtggga aggggtcaca ggaagggcct tgtaggccca gggagccgcg 20160
tggaaatgtg agtagcttag cgtacgccag agaggtggga aatggggccg ggccactgcg 20220
ggaggggtgt tcagaatcgg ccactttct cttgtttaaa atatctgctc tgctggggtc 20280
cccttactca gggtgtagct attctaatta tccttattgt tggaggcttc aaaatcaatc 20340
agtcgggccc cacatgtacc acttctgttt atttatatcg gttggttggt tgatttacag 20400
ggtctccctc tgtctcccag gctggaaccc gggtgtttga agctgcagtg agctatgatc 20460
gcaccgcacg gtcatgtgat cgcacagtgg tgagatcatc gcccactgca gcttcaaaca 20520
cccaggttca agggatcctc ccacctcagc ctcctgagta gctgggacca caggcacaga 20580
ccaccacacc gggctaatta aaaaaaattt ttttttttag aggtagagta ttgctgtttt 20640
acccaggctg gtctcaaact ggcctcaagt gatcctccca ccttggtttc ccaaagtgct 20700
ggcattacaa gtgtcggcca ctgagctagg ttcttggata ctctctttaa gaaaatgcct 20760
cttcttggaa gttgcagtga gccaagatcg tgccactgca ctccagcctg ggcgacaaag 20820
cgagactcca tctcaaaaaa aaaaaagaaa aagaaaatgc ctcttcaagt cttttgatag 20880
tttttctgat ggaggatttg tcttttttcct gttgatttgt agggctttct tacattctcg 20940
atcaaagtct tttgcccaca tataagtaca ttacaaatat cctctcccac actgcagctt 21000
gcctttcctt cctttaatgg cgttttttat tttattttat tttatttat tttattttat 21060
tttttgagac agattctcac tctgttgccc aggctggagt acagtggcgt gatctcggct 21120
cactgcaacc tccgcctccc tagttcaagt gattctcctg cctcagcctc ccgagtaggt 21180
gagattacag gtgcgcgtca ccaggcccgg ctaattttg tatttttagt agagacgggg 21240
tttgccatg ttggtcaggc tgggctcaaa ctcctgactc aagtgatcct cccgcctcaa 21300
cctcccaaag tgctggaatt acgtgcatga gccactatgc ccgaccaaaa atgtcatttt 21360
aaatacaaaa atgtattttc atcttcaggt ttgtcacttg attcagaatg gccatatttc 21420
aaatctccac cacccagagg aaggtgtttc ccaccttgct gtattcagcc agcgttttc 21480
acctccttca gcttgaaagt cctgggcaag gggcactggg cttgttcctc tggccctgca 21540
```

POR.ST25.txt

```
ggcttggggc aagggtcaga aggaaacaga agaacggaaa atggggcggg gtgagggggtg  21600
gtgcggggggt agaaggccag aagtggacag tggacagaag attgtgtcac cccagcctga  21660
tcctgaggct ggcgggcacc aggctgccct ctggagggga tggaccaagg tgagggctgt  21720
agtcccagga acgcagagaa gggaaaaatg aaatgtggcc cagtgcaggg gaagcagagg  21780
gggtctttga atcccagggg agctgaccct tgcgcaggac cttgagggca gagcggggggt  21840
cagccaggtg aggggactgt ggcaggaagg ctctgggtgg aagcgatggc gtgcacagga  21900
ttctgaggcc accatgcagg gaggggggtgt gaattaggca tcaagcttgg tttgtggagg  21960
ggagtggcag gtgacgctgg agacagactc tgagggggact catgtgcccg gatcccatgg  22020
gtcgtgggag ctgttgaaga gcggagcagg gaagtggcag atgtgagagg caggggctgg  22080
agaaggaaga ggtgaagagg tggaaggtca ggctgaggtc aggccgtggg gacctgggct  22140
gtgggtgggg aggaagaggg aagtggaaac cttggctctt ccttgaggga aacaagccca  22200
tatctatctg ggtgcaggag aggtcttggc tttgcaaagc actggtgtcg caagaggctt  22260
tactgtaggg gaaatgggaa gggtgagctg gtgcgggagg cctgggagcc ctggtgttgg  22320
attagaggcc agctgctggc ttagcccctg gggagtggca cgggacaaag ctggaatgtc  22380
ccctcccctgt gctgccacag catcccatga aacctggcatc taaggacacc cacgtcccgt  22440
gacacctgtg tcccatgaca cctgcctccc acgctcattg cacacttttg tcttgcagga  22500
cctcctctgt cagagagagc agctttgtgg aaaagatgaa gaaaacggtg agtttcctgc  22560
atgtctttac tcttctcagg aagcctcggc cccgatgggc atgtaatcaa gtctagcaga  22620
cggctggcgt caccrcatag aggatgtccc ggtggccagg agagggaacg cagcccggct  22680
cgcttgggag tcattcttgc ctaaccaaag ccctgaagca gctctgctgt tgctgggaca  22740
gtttgtttcc ttgaatcttt acttttgaa caaaccaaag tctaagaaca gaccctccct  22800
ctaactcacg aagcttctcc cgtctggccc tggcctgtcc aggagcagag gtaaatctca  22860
cagagctgcg gtcgtggcac cactgccatt ttctctccgc tttgttttttg ttactttttt  22920
ttttttgaga tggaatctcg ctctgtcacc caggctggag tgcagtggcg caatctcggt  22980
tcactgcaag ctccacctcc cgggttcaag ccattctcgt gcctcagcct cccgagtagc  23040
tgggattaca ggtgcgtgcc accacgccca gctaattttt gtattttaa tagagatggg  23100
atttcaccat gttggccaga ctggtcttga actcctgacc tcaggtgatc cgcccgcctc  23160
agcctcccaa agtgctggga ttacaggcgt gaaccaccgc acccagccaa tttaactatt  23220
taattgtagg catcagtact acatttacat agttaaaata aaaacatcca ctattcagtg  23280
gaaagctccc ctctgccacc cactcatccc cactgttctt attatttatg gatccttcta  23340
gagttcctgt atgaacatac aagtaaaaca aacagagact cttatttttta atttaaaaaa  23400
acctttattt tattttagag acagggtctc actgtgttgc ccaggctgga gtgcagcggc  23460
gccatcatag ctcactgcag cctcaaactc ctagactcaa gccatcctcc taccacagcc  23520
tcccaagtag ctgggactac aggtgcacac caccacgcct ggctaatttt ttaatttatt  23580
cttcatagga acaggttctc gctatgttgc ccaggctggt cttgaactcc tggcctcaag  23640
caatccttct gcctcccct cccaaagtgc tgggactgta ggtgcgagcc attgcatcca  23700
gcctacactt ttcttaaaac acaaaacgtg gcatactatg tgcactgttc cagaccttgc  23760
tttttttttt tttttttttt tttttgagac agagtctcgc tctgtcgccc aggctggagt  23820
gcaatggcac ggtcttggct cactgcaacc ttcacctcct cggttcaagt gattctcccg  23880
catcagcctc cggagtagtt ggggttagac gcctgccacc gtgcctggct aattttttgta  23940
tttttagtag agacggggtt tcaccatgtt gaccaggctg gtctcaaact gctgacctcg  24000
tgatccatct gcctcggtct cccaaagtgc tgggattaca ggcgtgagcc actgcgcccg  24060
gctgacctg ctttaaaaaa aaaagaacac aaaacaactt gatatgtctc aaagatcctt  24120
ccatatcaat acaaagatac tgccatcatc ttagctgtac catgattcca ttatatggcc  24180
atctggattc actgtaacgc acttaacct gtcctattaa agatgggcac tggggctgtt  24240
gctactcttt ggctcttacc aacaatgtca ctgtgaatat gtagctcggt ttgcacattt  24300
caaatacaaa gcagtcccag aagtggaatt gctggattgg aaggtagatg tgtttgcaat  24360
cgtgatagaa gctacaaatc gtccccgtag gagtcacacc gtggcacgcg ctcaagaata  24420
gcgacgtgtg agaaaactgt tccccacagc cttgccaaca gtttgttatt aaagtttaga  24480
tttttgtcag ttttctaggc ttaaaaaaaa aactcagtat agttgtaatt tacatttttta  24540
tgaatgagat tcagtagatt ttcacccacc taggtctaga tggcttcgtt tcttttttctt  24600
ttttcttttc tttttttttt tttttttttt gagacagggt ctcgctttgt tgcccaggct  24660
ggagtgcagt ggtgtgatct cagctcactg caatctccac ctcccgggtt caagcaattc  24720
tcatgcctca ttctcctgag tagctgggat tacaggtgcg tgctaccatg cctggctaac  24780
ttttgtattt tcagtagaga cagggttttg ccatgttggc caggctggtc tcgaactcct  24840
gacctcaggt catccactca acttggcctc ccaaagttct·ggaattacag gcctgaacca  24900
ccgcacccag cctctttttt cttttttaa tagattcctt cttcacatct tttgcctgct  24960
tttctgttgg gattgatctt tctccttgaa attattaggg agatgaatct cttgtctgtg  25020
atataagttg aaaatatctt· gcccgttaaa gagtttgtct tttagtgaga ttccatctct  25080
taaaaaaaag tttgtctttc aacctagctg atcgcatttt tatttgtttt gctgtgtaaa  25140
agttacagag tcacctatat gattctttct tttcgtgact tctggtctta gagtcataat  25200
ttgaaggtct tccccatgct aaagagataa agatattcct gtattttttt tttctagtac  25260
·ttttagggtt ttgttttttca tattcaaaat ttaaaatctg atctattggg aatttatcct  25320
gatgtaagat gtgaggtatg aatccaactc tacattttac atagctgtct aaacaccatg  25380
tattagagtc catctttccc ccactggttt gaaatgccac ctttatcata tacttaattc  25440
ctgggctggg cgcagtggct catgcctgta atcccaacac tttgggaggc cgagatgggc  25500
aggtcgcttg agcccaggag tttgagacca gcctgggcaa catggcgaaa ccccatctcc  25560
acaaaaaata gaaaaattag tgggacacag tggtgcgcac ctatagtccc cgctacctga  25620
```

Page 9

POR.ST25.txt

```
gaggctgaag tgggaggata gcttgagctt gggaggcaca gattgcagtg accactggac 25680
tcaaacctgg gcaacagcta caccctgtgt caaataataa taataatagc tgggtctctg 25740
ggttcttttc actgggcgca acgtgttcaa ggttcaccca cactgtagca tatgttagta 25800
cttccttttt attgccaaat aatgttttaa tatatggatt cttacttatt tttccgtgac 25860
ttttaggatc aatttgttga ggaaaaaaaa gctgttcaga tttttttttt tttttttggt 25920
agagactggg tctcactctg ttgcccaggc tgttctcaaa ctcctgggct taagcagtcc 25980
tcctggtgtg gcctcccagg gtactgggat tacaggtgtg agccactgca cccagctggt 26040
tcagatttct attacatcat gccacatttc tctagtgaca taaatagagc tgccattttt 26100
gtgggtcgag actccctgtc cataatcggg gtgcttttcc attggcgtga gtcctttgtg 26160
tgtagttcag tagcttctga cacatctctt acaaagttta tttctaggta ttttatatat 26220
tttgatgctc ctgtaaatat ttagctgcta tcttaggtca tcttttgttc tcttaacctg 26280
ataatcatga gcactttctc ttctattaag tatttttacg gtgattattt taaaatatcc 26340
ttttgggcga aaaccctgtc tctactaaaa atacaaaaat tagccgggcg tggtggcatg 26400
cacctgtagt cccaggtact cgggaggctg atgcaggaga atcacttgaa cccgggaggc 26460
ggaggttgca ttgagccaag atcgcaccac tgcactccag cctggccaac agagcaagat 26520
tccatctaaa aaaaaaaaaa atccttttgg ccgggtgcag tggctcatgt ctgtaatccc 26580
agcactttgg gaggctgaga cgggcggatc acctgaggtc aggagctttg agaccagctg 26640
gaccaacatg gtgaaaccct gtctctacta aaaacacaaa aaattagcca ggcatggtga 26700
cggacacctg taatcccagc tattcaggag gctgaggcag gaaaatcgct tgaacccagg 26760
aggcagaggt ttcagtgagc cgagatcgca ccactgcact ccagcctagg cgacaagagc 26820
aaaactctgt ctcaagaaaa aaaaaaacaa ttttttaatga tttaacaatg atttgtggaa 26880
aattcaggaa ctttaaaaag tataagaaca tttattgtaa tccctaagtt tttttttggtg 26940
gtggtttttgt tttgagactc agtctcactc ctgtcaccca ggctggagtg cagtggctca 27000
atcacagctc acttcagcct caacctccca ggctcaagtg atcctcccac ctcagcctcc 27060
caagtagcta gaaccatagg catgcgccac cacacccaga taaatttttt gtatttttag 27120
tagagacagg gtttttgccat gttgcccagg ctggtctcaa actcctgggc tcaagtcatc 27180
ctcttacctc ggcctcccaa agtactagga ttacagaagt gagccacggc acccggccgc 27240
gtaagtttta ttttgtatga ggtgggattc agtgtgaaaa ctagagacta ctttaggcat 27300
tttaatctgg aagggatttc acacagggaa gggggtactt ccaggttact ggagggggttc 27360
aaggagcagg ttggtctcta agatatcaca gaagcaacct gcaaggtagc tgctctgttc 27420
gaatgcagcc tgaaactgtt gagttcagga cacaccacgt gtagctgtga tccaggataa 27480
gaagctactt ctacagctgc ctctgaacat cacataagcc agtagccagg actaggaact 27540
tgcagtctgg ctaccaaaac tattagaaat acctcaggtg gctgggtgcg gtggctcacg 27600
cccgtaatcc cagcactttg ggaggccaag gtgggcagat cacctgaggt caggagttcg 27660
agaccagcct ggccaacgtg ttgaaaccca gtctctacta aaaattcaaa aaaattagcc 27720
tagcatagtg gtgggtgcat gtaatcccag ctactcagga ggctgaggca cgagaattgc 27780
ttgaacccag gaggcagagg ttgcagtgaa ctgagatcgt gccattacac tccaatctgg 27840
gcgtcagagc gagactccgt ctcacacaca aaaataaaaa aaagaaaaga aaaaaagaaa 27900
aatgccttag gcactgaaga ccaggcactg cagagaacct cacatgtgca tgaggggcca 27960
ggaagccttg gcctcccttc caccccttcag atcttgcaaa gcacgttcaa ttggctgaac 28020
cgaattggct gcaactccag ccccaacggg ctctggggaa taggatttca ggcttttcaa 28080
cagctgctgg aggaggaggg tggactgacg agtatgtgag ccagccccta ggacctttcc 28140
ctgttgtaaa attttataat gaatatgtat ttaaccagca aaacacatcc agaatgcccc 28200
atttagggac cctcggtata tttcttgcca gtctgttttc tttgtgcctg tgggtataga 28260
aagaaaatac atatgtatgt tttgatatac atatctatta tttatataac acaattggat 28320
tgtatagcca gccctgcata tccgtgagtt cggagttcaa gaattcaact gcagactgaa 28380
aatatttgga gaaaaaatgg attgttgtgt ctgtcctgaa cacatactgt catcttactg 28440
ttccctaaac aatacagcat agcaactatt tacagagaat ttacattaca ttaagtatta 28500
cattaataag taatctagag atgatttaaa gtatgttgga ggatgtgtgt aggctacaca 28560
caaataccac atcattttat agcagcgact tgagcatcct tggatttttg ttttttgggt 28620
ttttttgtga tggagttttg ctcttgtagc ccaggctgga gtgcaatggc gcgatctcgg 28680
ctcaccacaa cctctgcctc ccaggtccaa gcgattctcc tgcctcagcc tcccaagtag 28740
ctgggattac aggcatgtgc caccatgccc ggctaatttt gtatttttag tagagacagg 28800
gtttttctcca tattggtcag ctggtcttg aactctcgac ctcaagtcat ccacccgcct 28860
tggcctccca aagtgctggg attataggcg tgagccactg gaggtggagt cttgctctgt 28920
cacccaggct ggagtgcagt ggcgtgttct cagttcactg caacctccac ctcctgggtt 28980
caagctgttt tcctgcctca gcctccctag taactgggat tacaggtgca cgccaccatg 29040
cccggctaat ttttgtagtt ttagtagaga tggagtttca ccatgttggt caggctggtc 29100
ttgaactcct gacctcaggt aatctgccag cctcagcctc ccaaagtgct gggatcacag 29160
gtgtgagcca ccacgcccgg ctgcatcctt ggatgtaagt atctatggga ggtcctggaa 29220
ccattccccc acggataggg gctatacatt tggttacttt ttttaacttt gtttttttccc 29280
cgtgtcatga aaatgtcaga agtgtgatct tgagcagccg cgtgtgtgag attgccttgg 29340
tgacctttgc cctcctttgc cacagtggct gtgacagtga gaagcaagtc ccagaggaac 29400
ttagaaggga ctcaaagcca ggaaggaaag ggggcggcct ggagggcccc cgcctgccag 29460
gcctgcccag tgggtgttca ccggagccgt ggctgaggtc tgtggccctc accaaccctg 29520
tgtctgcctt ccttagggga ggaacatcat cgtgttctac ggctcccaga cggggaytgc 29580
agaggagttt gccaaccgcc tgtccaagga cgcccaccgc tacgggatgc gaggcatgtc 29640
agcggaccct gaggagtatg acctggtaag ctgccaccgc gtgctggccc cagatggagg 29700
```

Page 10

POR.ST25.txt

```
cagtgggtag gacagggagc aggtctgtag ggcgcccctc agcaggggga ggcyggcagg 29760
gagtggggtc ctgggaagac gtcctcggaa gttgccttcc cgtgagggtc attgccatcc 29820
ctgcccgggc tgactgacgg aagggccatt cctgcagttg cagccacgtg ccaggcatcg 29880
tttccccagg actggccctc ccctgagtgc cggctgccct cctggcggcc gccacatctc 29940
ccaaaccaaa cccaccctcc ccagctgctc cactcccctc tccttcgagg tcttggtgac 30000
gggcactacc ccacagccct ctccaggctg catttgcagt tgatgaacag tttagttaga 30060
tatcagtgtg caccggggct gtgcggtgcc tgggagcttt ggaatgaggt tcccagggtg 30120
gccccagacc tgcatcagct caggaggctg cattctgatg gctcttttgt gccaagcttc 30180
actttcctct agaaacacat cctctccgct ccagcctaac acgggtgacc ttgtcctgcc 30240
cacctggcac aaatgcccat gccccagccc ctccgtgttg ttacttctct ctgatcccac 30300
gacactcaga catccctggc ctggtgccac cctgggcagg acctggcctt ccccatctgg 30360
tgcgggttga accttgaaca ggctcagtca tggccggggc gcggtcctgt ccctgtttct 30420
gcaggcygac ctgagcagcc tgccagagat cgacaacgcc ctggtggttt tctgcatggc 30480
cacctacggt gagggagacc ccaccgacaa tgcccaggac ttctacgact ggctgcagga 30540
gacagacgtg gatctctctg gggtcaagtt cgcggtgagt cacccagaga ctgctatggg 30600
ctcccggtgg cctgcggtgc ctccctgggg actccagatc catgtatctg aaaggcagcc 30660
ctccagaccc ccaccctgtc ctcagcagcc agggcaggcc acttgtgaga ccctctcctc 30720
tgccccagac cccagcacta cgagaatgtc ccctccctgt ccccagcccc agtcggctcc 30780
agggggcatt gggtgctgga gactaaaggc aagaaaggtc tggctggacg actgagacct 30840
gcctggcctc ggagagctgg agccccagcc cggtggggag ggtgggctgg ccccgcttcc 30900
ctgtgggttg aggctggcag tggacggggc aggctgtggg ctgcaggtca accagatgaa 30960
gcctcttcag tggcccagtg ttccttgcag tgcggacgcc ctggtggaac ggaggcctgc 31020
aggtgttgcc agcagctcag ccagtgctgt ggggcctccc gccctgcccc catggcccct 31080
cccactggtc aggtcgaggg ccaggcctca gagcggcccc tgtgtccacg caggtgtttg 31140
gtcttgggaa caagacctac gagcacttca atgccatggg caagtacgtg gacaagcggc 31200
tggagcagct cggcgcccag cgcatctttg agctggggtt gggcgacgac gatgggaagt 31260
gagtgcccac cctgccacca tgatcagcgc ggcgggctta ggcaggggcc gtggaacgtg 31320
aggggcgcgc acaccattgt gtcagctgag actcagcgac acgcacctcc aacacagagg 31380
gaaggggctc tcctgcctct gccctgcccc tgccagtttt gcttttcggc ttgcccaact 31440
ccctggagcc ttcctgatgc tctgggttta tgtcgctggg tgccccaggg tgcacagtcc 31500
tgagctttgg ggatggggtg gggtcggggc rtgcctggca ccaggtaccg ttgccacatg 31560
ggcctccct gagccgctcc ccctctcctc tcctcggccc agcttggagg aggacttcat 31620
cacctggcga gagcagttct ggcyggcygt gtgtgaacac tttggggtgg aagccactgg 31680
cgaggagtcc aggtgagcaa gtgcccgcag gtgcggtggg tggcctgggc gggtcctgtg 31740
ccgagggcag ccaccctgga acaagggctg gcagtgggtc gcagcaaggt tagaagacac 31800
tccgtcatag ggtcgaggag ggaccttggt cccagccaag gactcactct gccacgttgc 31860
tctgcactgc cctgggggcag cggggtgcat cccacctctc gacaaggaca catcgcgtcg 31920
ggctctgtgg ctaggttcaa cttttggcaaa agggctcatt tccttaaaat ctgcctccac 31980
agacttggcc aaaaacataa cgttcctctc tgtgcccttg atggcccctg ggtgctgccc 32040
gggcttcctt accttctccc agatggaagc ctgcccagcc ctgccccggc ttctgggcgt 32100
ctggcccccg gcagctccac gccgcctccc tccttgctcc ctcgcctgcc ctccttgtgc 32160
atctgcagca ggggctcccc tgcttcttgt cgtatgtacc tgggacctca ccccaaaggc 32220
catgcacggt ctcccctgta gtccaacccc tccctctgta gactgaccc tgccgcttcc 32280
cggcctcacc cttggtctcc cctttccagc attcgccagt acgagcttgt ggtccacacc 32340
gacatagatg cggccaaggt gtacatgggg gagatgggcc ggctgaagag ctacgagaac 32400
cagaagccgt gagtggaggg agcgtggctt ggggcagacg gctctrtggc cactggtgca 32460
ccccaggctc agtctgcygt gtatccccat atcccacag ggcccttctc accagacccc 32520
gtgccccgag tgggtgtgag tgtccacgac ctgtccacgg cccggcccca ggagactggc 32580
actcgcccag tttgtactga gctcacttag ccaccttact ctgcactgcc cagggcagcg 32640
gggcacatcc cacttctcga ctaggacagt gcattggcct ctggtgctag gatcctcttt 32700
ggcagaaggg ctcattttct tagaatctat ctccatagat agacttgggg gagaaacctg 32760
gcgttcctct ctgtgccctt gatggcccct ggttgctgcc cagattcctg cccgagccac 32820
ccagggaagt cctcgtgcca ctgcatccaa gaccagctgc cctgctttct gtaggcctgc 32880
cccagcctgg ctggagcgag aagccctgca tgtcggtggc tttccggagt gcaagtttct 32940
ctcttgtcca cggtgtagtc caggacgtgt gggggtgggg gagctgtgct cctcatttgg 33000
ggaccagcca ccttccagcc tgtggcaccg tcagcttggg cctcacagtt cttttagggg 33060
ccagcctcag tttccacatc tgtaaacaga cactgatgac ccagctctgc ccatgtggtg 33120
ggactataga gagaactgca ttagggctgg gcgaggtggc tcacagctct aatcccagca 33180
ctctgggagg ccgaggtggg tgaatcacct gaggtcagga gttcaagacc agcctggcca 33240
acatggtgaa accccgtctc tagtaaaaat ataaaagtta accaggcatg gtggcgaatg 33300
cctgtaattc cagctgctca ggaggctgag gcaggagaat tgcttgaacc tgggaggcag 33360
aggttgcagt gagccgagat cgcgccgctg cactccagcc tgggtggcag agcgaaactc 33420
tgtctcaaaa aaaaaaaga gaactgcatt ggaccaggct gggagagccc ttgatgtaac 33480
cggtgagatt tcctcatgga gatctctgag attccctgtg ctttgtgcaa ccagaagsgt 33540
ccttgggagc rgagactcag atcaaagccg yggccgtcca ctgtgcttct ctcctcccca 33600
cccagcccct ttgatgccaa gaatccgttc ctggctgcag tcaccaccaa ccggaagctg 33660
aaccagggaa ccgagcgcca cctcatgcac ctggaattgg acatctcgga ctccaaaatc 33720
aggtaccagc tgccactgtc accccctgaa ccctcactct gggcctcctg acctggggca 33780
```

POR.ST25.txt

```
gggccagcct tccgcccctc ccgagcctca catctccctc caggtatgaa tctggggacc 33840
acgtggctgt gtacccagcc aacgactctg ctctcgtcaa ccagctgggc aaaatcctgg 33900
gtgccgacct ggacgtcgtc atgtccctga acaacctgga tggtgagtgc cacagtcagg 33960
gcgccctgcc gggctcaggc agccgcggga ttgggcctgt aggaaggccc tgggttgagc 34020
ttctgcttag gcctgaagcc ccggtgcctg ggaggccctt gcaccgagac tccacggtta 34080
caggatccca agcaaacggg aggcggggtg gccctagggg tctagccctc tctgtcgggg 34140
ttccccctac cccgtcactg tcatagtcct ttaaggpagt gaggtgctga ggcctggtgg 34200
cagaggcagc cctggctccc ccatggccac tgtgtcctgc tgggaaggag ggcctggctc 34260
cacgacccac ctctgccggc ctggggctgc ccccacctcc tcactgaagt caggagtcag 34320
cagccctccc aggcccccaa gggtgcacag tggtgccgag tggcagtagc cattacgcgg 34380
ggctgcctgg gcctggtggg gctgcccagc ctgagcctcc cgctgtgaag ccctcggacc 34440
ccactggtca ccaacctggg ccgagcccac ctcgccccac ccctgcttgc cggtcctcag 34500
ctgccatgcc agggctgccc ttactgtcgg cttccacagc ccgcctgtag ggaagcctgg 34560
gcggggctgt gtcagaccgt gtagtgtgcg gtgaaagcac agcgggtgcg ttagtgtgca 34620
ggggctcccc cgcacctccc ctagcagggt cctgcctctg atgaggactt cctgtctggt 34680
tggaggcccc agacttggcc ccagggccag ggaggcatca gagagcatag gccttgtttc 34740
cagcaccagc tggggcacct gttgcygcag agctggccca aggtgtcacc ccctcyygcc 34800
gcagccaccc atccccagga ggcggccgcc taccccaagt cctgcctgts tcttccctgc 34860
agaggagtcc aacaagaagc acccattccc gtgccctacg tcctaccgca cggccctcac 34920
ctactacctg gacatcacca acccgccgcg taccaacgtg ctgtacgagc tggcgcagta 34980
cgcctcggag ccctcggagc aggagctgct gcgcaagatg gcctcctcct ccggcgaggg 35040
caaggtgcgc cccctcagcc cccgcaacct ccgcccgccg accccgccgt tttccgagct 35100
ccgtgggccc caatcagccc catctcaccc ccgtgtctct taggagctgt acctgagctg 35160
ggtggtggag gcccggaggc acatcctggc catcctgcag gactgcccgt ccctgcggcc 35220
ccccatcgac cacctgtgtg agctgctgcc gcgcctgcag gcccgctact actccatcgc 35280
ctcatcctcc aaggtgaggg ccggcactgc cctgccagcc acacrctgga ggcccagccc 35340
tgctcacagc aggcagagtg caaggcggca caggagctcc gagatctgag ccctgagctc 35400
cagttccagc cccagcgcag ctccaaatgc ctccccaggc tgtggactca gtcgggctgg 35460
cttgtgagat tctcagcatc tgtccagccc cggtccccag aaccagtccg ggaagccgct 35520
ggggaggggg cctctgaggt ttgggtgcca ggtgccctgg aagagggccct gggtgagtgg 35580
ggctggcctg cagaacggga cttggggccg gggctgggca agggcctcgg yktggcggtg 35640
gagctcacac ggccctcccc acaggtccac cccaactctg tgcacatctg tgcggtggtt 35700
gtggagtacg agaccaaggc yggccgcatc aacaagggcg tggccaccaa ctggctgcgg 35760
gccaaggagc ctgycgggga gaacggcggc cgtgcgctgg tgcccatgtt cgtgcgcaag 35820
tcccagttac gcctgccctt caaggccacc acgcctgtca tcatggtggg ccccggcacc 35880
gggtggcacc ctttcatccag cttcatccag gagcgggcct ggctgcgaca gcagggtgag 35940
tggggtccca tggggggagg ggggtgacga ctgggagccc cgcgctcacc cyggcccctg 36000
ccacgcaggc aaggaggtgg gggagacgct gctgtactac ggctgccgcc gctsggatga 36060
ggactacctg taccgggagg agctggcgca gttccacagg gacggtgcgc tcacccagct 36120
caacgtggcc ttctcccggg agcagtccca caaggtgaga cgggcgggca cccacgaagg 36180
tgggcatgag gctggcaggg ccacagscac agtgcccccc tcacagcacc acccttggcc 36240
ccaggtctac gtccagcacc tgctaaagca agaccgagag cacctgtgga agttgatcga 36300
aggcggtgcc cacatctacg tctgtgggtr agtgagtggg gtcactgaaa tagggggcag 36360
ggaggacaag gccctgcctg ccacagttgg cccagccccc agcacccct cttcctgccc 36420
agggatgcac ggaacatggc cagggatgtg cagaacacct tctacgacat cgtggctgag 36480
ctcggggcca tggagcacgc gcaggcggtg gactacatca agaaactgat gaccaagggc 36540
cgctactccc tggacgtgtg gagctagggg cctgcctgcc ccacccaccc cacagactcc 36600
ggcctgtaat cagctcycct ggctccctcc cgtagtctcc tgggtgtgtt tggcttggcc 36660
ttggcatggg cgcaggccca gtgacaaaga ctcctctggg cctggggtgc atcctcctca 36720
gcccccaggc caggtgaggt ccaccggccc ctggcagcac agcccagggc ctgcatgggg 36780
gcaccgggct ccatgcctct ggagcctctg gccctcggtg gctgcacaga agggctcttt 36840
ctctctgctg agctggccca gcccctccac gtgatttcca gtgagtgtaa ataattttaa 36900
ataacctctg gccttggaa taaagttctg ttttctgtat ttgcctggta ttgtgtgagt 36960
agatctggga cctccacctg catcaactta actaactcag accccaggaa cccatgtggt 37020
ggggccaccc agccctcccc tcccccagga gaacacacac gctcaggcca cctctgggcc 37080
tctctttatt gagggcactg ggcccaggtc ttccttcagg gcccacagcg cccataaaac 37140
ccaagggaga atagaagaga cccctgata cacgcacact cgaggggcgc ctcccatccc 37200
ctcccacaac acacaggaca gaagccccte tgggcggca ggggaaggcc cagcctcaat 37260
ccttcttgct cccgtgccgc tgactgtgaa acttgtggtg cacaaccctc agggtggtga 37320
agaaattgcc gaggaaaagg aggaggaagg gaaagccgca cataagcacc tgccggagga 37380
ataggtgag ggctggacat gggcctggcc ccccatccat cctgtccagg gaccccggct 37440
cacctgccac tccttgcact gagggtcctg ggccaggttg aacaacgtca gcgcgttaaa 37500
aagctgccag aactaagcag ggaggaggca ttttactcag aagacccagt ccctgctgga 37560
gcccgaaacc ggcgcagagg acggacagac agacaggcag ggacacagag gcacgactta 37620
cgtgtccaaa gaaaagaaaa ggcagcagga aggtgaggcc ccgccacatc caggactgga 37680
agccctctgc ggggaggaag gtcagggcac aggacggcca gggacagagg ggacgctacc 37740
actcagaccc gggatgggag gaaaggggaa tgtccagatg ccagggccac tcccgactca 37800
cccacagtga ggtccatggt gtgccgctcg cccagcgccc gcaggcggta gaggcagccg 37860
```

Page 12

POR.ST25.txt

```
ctctggtagt agtactggag aaaactgcacg aagcctgggc cgggcggagg acagaggagc    37920
agggctgagc cccgggaggg gttccctgcc cctgcccccc accacgggtg cctccagggc    37980
tcagcacaga catctgggac actcactctg gtacatggaa aaggagagga attggttccg    38040
gaatttctgg tacatgagac cgtcgggcct aaggtaaaaa gtggagggca gtgtggggac    38100
ccttggggat gccgtgtatc ccctccttgt cccagctcct gcccctgccg gggcccagcc    38160
actcaccacg tcagcatgac tcccgacagg aaggtggaca cgtagtgatg gaacacccac    38220
cagcctttga tcctggagca gagagccacc atcacccca gcgcctgttc ctgcttcccg    38280
gagggtcctg gcaccccatt ccatgctccc ctccctcagg gcccgccctg cccaccggga    38340
gccgttgttg atgaggatgc tctcccggat ggtcagggtg cagtagtacc agaccagcag    38400
gaagttgaag gcagcatctg tcaccctgcg gagggagggg accgagggtt ggtactgcag    38460
cgactgggga tggggtgggt cctcggccgg ggtgggacgc ccacctggag ttgagcagga    38520
agcggcaagt gaaggagatg aggatgagga tgatggtgag gtagagcttg aacttctcat    38580
actcgtcctt gtaggcaaac ctggggggcg caggccggtg agacgggtgg ggtgctggtg    38640
gggcccatgt gtgcccccac cccgggaggg cagctgagcc aagggtagtc ggatggagga    38700
gaaggccggg ttgggggtg gagggaagg ccaggtgggg tgggggtgg aggggaaaac    38760
caggttgggg ggaaagccgg gtggagggg ggagggaag gccgggtggg gggctggagg    38820
ggaaggccag gtgggggct ggaggggaag aaggccgggt ggggggctgg aggggaaggc    38880
cgggtggggg ggtggagggg aaggcgggga ggggtgtgg aggggaaggc gggaggggg    38940
tgtggagggg aaggccgggt gggtgtgtgg aggggaaggc cgggtgggtg tgtggagggg    39000
aaggccgggt gggtgtgtgg aggggaaggc cgggtggggg ggtggagggg aaggccgggt    39060
ggggtggggg gtggaggga aggccgggtg gggtggggg tggaggggaa ggccgggttg    39120
ggggtggag gggaaggccg ggttccgggg gaaggctggg tggaggggtg gaggttgagg    39180
gggggagggg ggtatggggg taggggtgga gggtggaggt tggggggttc aggctcggct    39240
gagttactta gcctgcttgc tcaggagcgt gacgtttacg ttcccca                  39287
```

```
<210>   2
<211>   2031
<212>   DNA
<213>   Homo sapiens

<400>   2
ggagactccc acgtggacac cagctccacc gtgtccgagg cggtggccga agaagtatct      60
cttttcagca tgacggacat gattctgttt tcgctcatcg tgggtctcct aacctactgg     120
ttcctcttca gaaagaaaaa agaagaagtc cccgagttca ccaaaattca gacattgacc     180
tcctctgtca gagagagcag ctttgtggaa aagtagaaga aaacggggag gaacatcatc     240
gtgttctacg gctcccagac ggggactgca gaggagtttg ccaaccgcct gtccaaggac     300
gcccaccgct acgggatgcg aggcatgtca gcggaccctg aggagtatga cctggccgac     360
ctgagcagcc tgccagagat cgacaacgcc ctggtggttt tctgcatggc cacctacggt     420
gagggagacc ccaccgacaa tgcccaggac ttctacgact ggctgcagga gacagacgtg     480
gatctctctg gggtcaagtt cgcggtgttt ggtcttggga acaagaccta cgagcacttc     540
aatgccatgg gcaagtacgt ggacaagcgg ctggagcagc tcggcgccca gcgcatcttt     600
gagctggggt tggggacagc cgatgggaac ttggaggagg acttcatcac ctggcgagag     660
cagttctggc cggccgtgtg tgaaacactt ggggtggaag ccactggcga ggagtccagc     720
attcgccagt acgagcttgt ggtccacacc gacatagatg cggccaaggt gtacatgggg     780
gagatggggc ggctgaagag ctacgagaac cagaagcccc cctttgatgc caagaatccg     840
ttcctggctg cagtcaccac caaccggaag ctgaaccagg gaaccgagcg ccacctcatg     900
cacctggaat tggacatctc ggactccaaa atcaggtatg aatctgggga ccacgtggct     960
gtgtacccag ccaacgactc tgctctcgtc aaccagctgg gcaaaatcct gggtgccgac    1020
ctggacgtcg tcatgtgccct gaacaacctg gatgaggagt ccaacaagaa gcaccattc    1080
ccgtgcccta cgtcctaccg cacggccctc acctactacc tggacatcac caacccgccg    1140
cgtaccaacg tgctgtacga gctggcgcag tacgcctcgg agccctcgga gcaggagctg    1200
ctgcgcaaga tggcctcctc ctccggcgag ggcaaggagc tgtacctgag ctgggtggtg    1260
gaggcccgga ggcacatcct ggccatcctg caggactgcc gtccctgcg gccccccatc    1320
gaccacctgt gtgagctgct gccgcgcctg caggcccgct actactccat cgcctcatcc    1380
tccaaggtcc accccaactc tgtgcacatc tgtgcggtgg ttgtggagta cgagaccaag    1440
gccggccgca tcaacaaggg cgtggccacc aactggctgc gggccaagga gcctgtcggg    1500
gagaacgcg gccgtgcgct ggtgcccatg ttcgtgcgga agtcccagtt acgcctgccc    1560
ttcaaggcca ccacgccgt catcatggtg gccccggca ccgggtggca ccctttcata    1620
ggcttcatcc aggagcgggc ctggctgcga cagcagggca aggaggtggg ggagacgctg    1680
ctgtactacg gctgccgccg ctcggatgag gactacctgt accgggagga gctggcgcag    1740
ttccacaggg acggtgcgct cacccagctc aacgtggcct tctcccggga gcagtcccac    1800
aaggtctacg tccagcacct gctaaagcaa gaccgagagc acctgtggaa gttgatcgaa    1860
ggcggtgccc acatctacgt ctgtggggat gcacggaaca tggccaggga tgtgcagaac    1920
accttctacg acatcgtggc tgagctcggg gccatggagc acgcgcaggc ggtggactac    1980
atcaagaaac tgatgaccaa gggccgctac tccctggacg tgtggagcta g             2031
```

POR.ST25.txt

```
<210>  3
<211>  676
<212>  PRT
<213>  Homo sapiens

<400>  3

Gly Asp Ser His Val Asp Thr Ser Ser Thr Val Ser Glu Ala Val Ala
1               5                   10                  15

Glu Glu Val Ser Leu Phe Ser Met Thr Asp Met Ile Leu Phe Ser Leu
                20                  25                  30

Ile Val Gly Leu Leu Thr Tyr Trp Phe Leu Phe Arg Lys Lys Lys Glu
            35                  40                  45

Glu Val Pro Glu Phe Thr Lys Ile Gln Thr Leu Thr Ser Ser Val Arg
        50                  55                  60

Glu Ser Ser Phe Val Glu Lys Met Lys Lys Thr Gly Arg Asn Ile Ile
65                  70                  75                  80

Val Phe Tyr Gly Ser Gln Thr Gly Thr Ala Glu Glu Phe Ala Asn Arg
                85                  90                  95

Leu Ser Lys Asp Ala His Arg Tyr Gly Met Arg Gly Met Ser Ala Asp
            100                 105                 110

Pro Glu Glu Tyr Asp Leu Ala Asp Leu Ser Ser Leu Pro Glu Ile Asp
        115                 120                 125

Asn Ala Leu Val Val Phe Cys Met Ala Thr Tyr Gly Glu Gly Asp Pro
    130                 135                 140

Thr Asp Asn Ala Gln Asp Phe Tyr Asp Trp Leu Gln Glu Thr Asp Val
145                 150                 155                 160

Asp Leu Ser Gly Val Lys Phe Ala Val Phe Gly Leu Gly Asn Lys Thr
                165                 170                 175

Tyr Glu His Phe Asn Ala Met Gly Lys Tyr Val Asp Lys Arg Leu Glu
            180                 185                 190

Gln Leu Gly Ala Gln Arg Ile Phe Glu Leu Gly Leu Gly Asp Asp Asp
            195                 200                 205

Gly Asn Leu Glu Glu Asp Phe Ile Thr Trp Arg Glu Gln Phe Trp Pro
    210                 215                 220

Ala Val Cys Glu His Phe Gly Val Glu Ala Thr Gly Glu Glu Ser Ser
225                 230                 235                 240

Ile Arg Gln Tyr Glu Leu Val Val His Thr Asp Ile Asp Ala Ala Lys
            245                 250                 255

Val Tyr Met Gly Glu Met Gly Arg Leu Lys Ser Tyr Glu Asn Gln Lys
            260                 265                 270

Pro Pro Phe Asp Ala Lys Asn Pro Phe Leu Ala Ala Val Thr Thr Asn
            275                 280                 285

Arg Lys Leu Asn Gln Gly Thr Glu Arg His Leu Met His Leu Glu Leu
    290                 295                 300

Asp Ile Ser Asp Ser Lys Ile Arg Tyr Glu Ser Gly Asp His Val Ala
305                 310                 315                 320

Val Tyr Pro Ala Asn Asp Ser Ala Leu Val Asn Gln Leu Gly Lys Ile
```
Page 14

POR.ST25.txt

```
                325                     330                    335

    Leu Gly Ala Asp Leu Asp Val Val Met Ser Leu Asn Asn Leu Asp Glu
                340             345              350

    Glu Ser Asn Lys Lys His Pro Phe Pro Cys Pro Thr Ser Tyr Arg Thr
                355             360              365

    Ala Leu Thr Tyr Tyr Leu Asp Ile Thr Asn Pro Pro Arg Thr Asn Val
        370              375              380

    Leu Tyr Glu Leu Ala Gln Tyr Ala Ser Glu Pro Ser Glu Gln Glu Leu
    385              390              395                    400

    Leu Arg Lys Met Ala Ser Ser Ser Gly Glu Gly Lys Glu Leu Tyr Leu
                405              410                   415

    Ser Trp Val Val Glu Ala Arg Arg His Ile Leu Ala Ile Leu Gln Asp
                420              425              430

    Cys Pro Ser Leu Arg Pro Pro Ile Asp His Leu Cys Glu Leu Leu Pro
            435              440              445

    Arg Leu Gln Ala Arg Tyr Tyr Ser Ile Ala Ser Ser Ser Lys Val His
        450              455              460

    Pro Asn Ser Val His Ile Cys Ala Val Val Val Glu Tyr Glu Thr Lys
    465              470              475                    480

    Ala Gly Arg Ile Asn Lys Gly Val Ala Thr Asn Trp Leu Arg Ala Lys
                485              490              495

    Glu Pro Val Gly Glu Asn Gly Gly Arg Ala Leu Val Pro Met Phe Val
                500              505              510

    Arg Lys Ser Gln Leu Arg Leu Pro Phe Lys Ala Thr Thr Pro Val Ile
                515              520              525

    Met Val Gly Pro Gly Thr Gly Trp His Pro Phe Ile Gly Phe Ile Gln
        530              535              540

    Glu Arg Ala Trp Leu Arg Gln Gln Gly Lys Glu Val Gly Glu Thr Leu
    545              550              555                    560

    Leu Tyr Tyr Gly Cys Arg Arg Ser Asp Glu Asp Tyr Leu Tyr Arg Glu
                565              570              575

    Glu Leu Ala Gln Phe His Arg Asp Gly Ala Leu Thr Gln Leu Asn Val
                580              585              590

    Ala Phe Ser Arg Glu Gln Ser His Lys Val Tyr Val Gln His Leu Leu
            595              600              605

    Lys Gln Asp Arg Glu His Leu Trp Lys Leu Ile Glu Gly Gly Ala His
        610              615              620

    Ile Tyr Val Cys Gly Asp Ala Arg Asn Met Ala Arg Asp Val Gln Asn
    625              630              635                    640

    Thr Phe Tyr Asp Ile Val Ala Glu Leu Gly Ala Met Glu His Ala Gln
                645              650              655

    Ala Val Asp Tyr Ile Lys Lys Leu Met Thr Lys Gly Arg Tyr Ser Leu
                660              665              670

    Asp Val Trp Ser
            675
```

Page 15

POR.ST25.txt

```
<210>   4
<211>   15
<212>   DNA
<213>   Homo sapiens

<400>   4
cgtcaccrca tagag                                                      15


<210>   5
<211>   15
<212>   DNA
<213>   Homo sapiens

<400>   5
acggggaytg cagag                                                      15


<210>   6
<211>   15
<212>   DNA
<213>   Homo sapiens

<400>   6
gggaggcygg caggg                                                      15


<210>   7
<211>   15
<212>   DNA
<213>   Homo sapiens

<400>   7
tgcaggcyga cctga                                                      15


<210>   8
<211>   15
<212>   DNA
<213>   Homo sapiens

<400>   8
tcggggcrtg cctgg                                                      15


<210>   9
<211>   15
<212>   DNA
<213>   Homo sapiens

<400>   9
ttctggcygg ccgtg                                                      15


<210>   10
<211>   15
<212>   DNA
<213>   Homo sapiens

<400>   10
ggccggcygt gtgtg                                                      15


<210>   11
<211>   15
<212>   DNA
<213>   Homo sapiens
```

POR.ST25.txt

·POR.ST25.txt

```
<400>   11
cggctctrtg gccac                                                        15


<210>   12
<211>   15
<212>   DNA
<213>   Homo sapiens

<400>   12
agtctgcygt gtatc                                                        15


<210>   13
<211>   15
<212>   DNA
<213>   Homo sapiens

<400>   13
ccagaagsgt ccttg                                                        15


<210>   14
<211>   15
<212>   DNA
<213>   Homo sapiens

<400>   14
tggagacrga gactc                                                        15


<210>   15
<211>   15
<212>   DNA
<213>   Homo sapiens

<400>   15
aaagcccygg ccgct                                                        15


<210>   16
<211>   15
<212>   DNA
<213>   Homo sapiens

<400>   16
ctgttgcygc agagc                                                        15


<210>   17
<211>   15
<212>   DNA
<213>   Homo sapiens

<400>   17
ccccctcytg ccgca                                                        15


<210>   18
<211>   15
<212>   DNA
<213>   Homo sapiens

<400>   18
cccctccygc cgcag                                                        15
```

POR.ST25.txt

```
<210>   19
<211>   15
<212>   DNA
<213>   Homo sapiens

<400>   19
tgcctgtstc ttccc                                                    15


<210>   20
<211>   15
<212>   DNA
<213>   Homo sapiens

<400>   20
gccacacrct ggagg                                                    15


<210>   21
<211>   15
<212>   DNA
<213>   Homo sapiens

<400>   21
gcctcggygt ggcgg                                                    15


<210>   22
<211>   15
<212>   DNA
<213>   Homo sapiens

<400>   22
cctcggtktg gcggt                                                    15


<210>   23
<211>   15
<212>   DNA
<213>   Homo sapiens

<400>   23
ccaaggcygg ccgca                                                    15


<210>   24
<211>   15
<212>   DNA
<213>   Homo sapiens

<400>   24
gagcctgycg gggag                                                    15


<210>   25
<211>   15
<212>   DNA
<213>   Homo sapiens

<400>   25
ctcacccygg cccct                                                    15


<210>   26
<211>   15
<212>   DNA
<213>   Homo sapiens
```

POR.ST25.txt

<400> 26
cgccgctsgg atgag   15

<210> 27
<211> 15
<212> DNA
<213> Homo sapiens

<400> 27
gccacagsca cagtg   15

<210> 28
<211> 15
<212> DNA
<213> Homo sapiens

<400> 28
tgtgggtrag tgagt   15

<210> 29
<211> 15
<212> DNA
<213> Homo sapiens

<400> 29
tcagctcycc tggct   15

<210> 30
<211> 15
<212> DNA
<213> Homo Sapiens

<400> 30
ggctggcgtc accrc   15

<210> 31
<211> 15
<212> DNA
<213> Homo sapiens

<400> 31
gacatcctct atgyg   15

<210> 32
<211> 15
<212> DNA
<213> Homo Sapiens

<400> 32
tcccagacgg ggayt   15

<210> 33
<211> 15
<212> DNA
<213> Homo sapiens

<400> 33
aaactcctct gcart   15

<210> 34

POR.ST25.txt

```
<211>   15
<212>   DNA
<213>   Homo Sapiens

<400>   34
agcagggggga ggcyg                                                          15


<210>   35
<211>   15
<212>   DNA
<213>   Homo sapiens

<400>   35
cccactccct gccrg                                                           15


<210>   36
<211>   15
<212>   DNA
<213>   Homo Sapiens

<400>   36
tgtttctgca ggcyg                                                           15


<210>   37
<211>   15
<212>   DNA
<213>   Homo sapiens

<400>   37
ggctgctcag gtcrg                                                           15


<210>   38
<211>   15
<212>   DNA
<213>   Homo Sapiens

<400>   38
gtggggtcgg ggcrt                                                           15


<210>   39
<211>   15
<212>   DNA
<213>   Homo sapiens

<400>   39
ctggtgccag gcayg                                                           15


<210>   40
<211>   15
<212>   DNA
<213>   Homo Sapiens

<400>   40
gagcagttct ggcyg                                                           15


<210>   41
<211>   15
<212>   DNA
<213>   Homo sapiens

<400>   41
```

POR.ST25.txt

ttcacacacg gccrg   15

<210> 42
<211> 15
<212> DNA
<213> Homo Sapiens

<400> 42
agttctggcc ggcyg   15

<210> 43
<211> 15
<212> DNA
<213> Homo sapiens

<400> 43
agtgttcaca cacrg   15

<210> 44
<211> 15
<212> DNA
<213> Homo Sapiens

<400> 44
ggcagacggc tctrt   15

<210> 45
<211> 15
<212> DNA
<213> Homo sapiens

<400> 45
gcaccagtgg ccaya   15

<210> 46
<211> 15
<212> DNA
<213> Homo Sapiens

<400> 46
aggctcagtc tgcyg   15

<210> 47
<211> 15
<212> DNA
<213> Homo sapiens

<400> 47
tatggggata cacrg   15

<210> 48
<211> 15
<212> DNA
<213> Homo Sapiens

<400> 48
gtgcaaccag aagsg   15

<210> 49
<211> 15

POR.ST25.txt

```
<212>  DNA
<213>  Homo sapiens

<400>  49
cgtctccaag gacsc                                    15


<210>  50
<211>  15
<212>  DNA
<213>  Homo Sapiens

<400>  50
cgtccttgga gacrg                                    15


<210>  51
<211>  15
<212>  DNA
<213>  Homo sapiens

<400>  51
tgatctgagt ctcyg                                    15


<210>  52
<211>  15
<212>  DNA
<213>  Homo Sapiens

<400>  52
cagatcaaag cccyg                                    15


<210>  53
<211>  15
<212>  DNA
<213>  Homo sapiens

<400>  53
acagtgagcg gccrg                                    15


<210>  54
<211>  15
<212>  DNA
<213>  Homo Sapiens

<400>  54
gggcacctgt tgcyg                                    15


<210>  55
<211>  15
<212>  DNA
<213>  Homo sapiens

<400>  55
gggccagctc tgcrg                                    15


<210>  56
<211>  15
<212>  DNA
<213>  Homo Sapiens

<400>  56
gtgtcacccc ctcyt                                    15
```

POR.ST25.txt

```
<210>  57
<211>  15
<212>  DNA
<213>  Homo sapiens

<400>  57
ggtggctgcg gcarg                                                    15


<210>  58
<211>  15
<212>  DNA
<213>  Homo Sapiens

<400>  58
tgtcacccc tccyg                                                     15


<210>  59
<211>  15
<212>  DNA
<213>  Homo sapiens

<400>  59
gggtggctgc ggcrg                                                    15


<210>  60
<211>  15
<212>  DNA
<213>  Homo Sapiens

<400>  60
aagtcctgcc tgtst                                                    15


<210>  61
<211>  15
<212>  DNA
<213>  Homo sapiens

<400>  61
tctgcaggga agasa                                                    15


<210>  62
<211>  15
<212>  DNA
<213>  Homo Sapiens

<400>  62
ctgccagcca cacrc                                                    15


<210>  63
<211>  15
<212>  DNA
<213>  Homo sapiens

<400>  63
gctgggcctc cagyg                                                    15


<210>  64
<211>  15
<212>  DNA
```

POR.ST25.txt

<213>   Homo Sapiens

<400>   64
gcaagggcct cggyg   15

<210>   65
<211>   15
<212>   DNA
<213>   Homo sapiens

<400>   65
gctccaccgc cacrc   15

<210>   66
<211>   15
<212>   DNA
<213>   Homo Sapiens

<400>   66
caagggcctc ggtkt   15

<210>   67
<211>   15
<212>   DNA
<213>   Homo sapiens

<400>   67
agctccaccg ccama   15

<210>   68
<211>   15
<212>   DNA
<213>   Homo Sapiens

<400>   68
acgagaccaa ggcyg   15

<210>   69
<211>   15
<212>   DNA
<213>   Homo sapiens

<400>   69
tgttgatgcg gccrg   15

<210>   70
<211>   15
<212>   DNA
<213>   Homo Sapiens

<400>   70
gccaaggagc ctgyc   15

<210>   71
<211>   15
<212>   DNA
<213>   Homo sapiens

<400>   71
gccgttctcc ccgrc   15

POR.ST25.txt

```
<210>    72
<211>    15
<212>    DNA
<213>    Homo Sapiens

<400>    72
cccgcgctca cccyg                                               15


<210>    73
<211>    15
<212>    DNA
<213>    Homo sapiens

<400>    73
cgtggcaggg gccrg                                              15


<210>    74
<211>    15
<212>    DNA
<213>    Homo Sapiens

<400>    74
ggctgccgcc gctsg                                              15


<210>    75
<211>    15
<212>    DNA
<213>    Homo sapiens

<400>    75
gtagtcctca tccsa                                             15


<210>    76
<211>    15
<212>    DNA
<213>    Homo Sapiens

<400>    76
ggcagggcca cagsc                                             15


<210>    77
<211>    15
<212>    DNA
<213>    Homo sapiens

<400>    77
gggggggcact gtgsc                                            15


<210>    78
<211>    15
<212>    DNA
<213>    Homo Sapiens

<400>    78
tacgtctgtg ggtra                                             15


<210>    79
<211>    15
<212>    DNA
<213>    Homo sapiens
```

Page 25

POR.ST25.txt

<400>   79
gaccccactc actya   15

<210>   80
<211>   15
<212>   DNA
<213>   Homo Sapiens

<400>   80
ctgtaatcag ctcyc   15

<210>   81
<211>   15
<212>   DNA
<213>   Homo sapiens

<400>   81
ggagggagcc aggrg   15

<210>   82
<211>   10
<212>   DNA
<213>   Homo sapiens

<400>   82
tggcgtcacc   10

<210>   83
<211>   10
<212>   DNA
<213>   Homo sapiens

<400>   83
atcctctatg   10

<210>   84
<211>   10
<212>   DNA
<213>   Homo sapiens

<400>   84
cagacgggga   10

<210>   85
<211>   10
<212>   DNA
<213>   Homo sapiens

<400>   85
ctcctctgca   10

<210>   86
<211>   10
<212>   DNA
<213>   Homo sapiens

<400>   86
aggggaggc   10

POR.ST25.txt

```
<210>   87
<211>   10
<212>   DNA
<213>   Homo sapiens

<400>   87
actccctgcc                                              10


<210>   88
<211>   10
<212>   DNA
<213>   Homo sapiens

<400>   88
ttctgcaggc                                              10


<210>   89
<211>   10
<212>   DNA
<213>   Homo sapiens

<400>   89
tgctcaggtc                                              10


<210>   90
<211>   10
<212>   DNA
<213>   Homo sapiens

<400>   90
gggtcggggc                                              10


<210>   91
<211>   10
<212>   DNA
<213>   Homo sapiens

<400>   91
gtgccaggca                                              10


<210>   92
<211>   10
<212>   DNA
<213>   Homo sapiens

<400>   92
cagttctggc                                              10


<210>   93
<211>   10
<212>   DNA
<213>   Homo sapiens

<400>   93
acacacggcc                                              10


<210>   94
<211>   10
<212>   DNA
<213>   Homo sapiens
```

POR.ST25.txt

<400> 94
tctggccggc 10

<210> 95
<211> 10
<212> DNA
<213> Homo sapiens

<400> 95
gttcacacac 10

<210> 96
<211> 10
<212> DNA
<213> Homo sapiens

<400> 96
agacggctct 10

<210> 97
<211> 10
<212> DNA
<213> Homo sapiens

<400> 97
ccagtggcca 10

<210> 98
<211> 10
<212> DNA
<213> Homo sapiens

<400> 98
ctcagtctgc 10

<210> 99
<211> 10
<212> DNA
<213> Homo sapiens

<400> 99
ggggatacac 10

<210> 100
<211> 10
<212> DNA
<213> Homo sapiens

<400> 100
caaccagaag 10

<210> 101
<211> 10
<212> DNA
<213> Homo sapiens

<400> 101
ctccaaggac 10

<210> 102

POR.ST25.txt

```
<211>   10
<212>   DNA
<213>   Homo sapiens

<400>   102
ccttggagac                                                                    10


<210>   103
<211>   10
<212>   DNA
<213>   Homo sapiens

<400>   103
tctgagtctc                                                                    10


<210>   104
<211>   10
<212>   DNA
<213>   Homo sapiens

<400>   104
atcaaagccc                                                                    10


<210>   105
<211>   10
<212>   DNA
<213>   Homo sapiens

<400>   105
gtgagcggcc                                                                    10


<210>   106
<211>   10
<212>   DNA
<213>   Homo sapiens

<400>   106
cacctgttgc                                                                    10


<210>   107
<211>   10
<212>   DNA
<213>   Homo sapiens

<400>   107
ccagctctgc                                                                    10


<210>   108
<211>   10
<212>   DNA
<213>   Homo sapiens

<400>   108
tcaccccctc                                                                    10


<210>   109
<211>   10
<212>   DNA
<213>   Homo sapiens

<400>   109
```

POR.ST25.txt

ggctgcggca   10

<210>   110
<211>   10
<212>   DNA
<213>   Homo sapiens

<400>   110
cacccctcc   10

<210>   111
<211>   10
<212>   DNA
<213>   Homo sapiens

<400>   111
tggctgcggc   10

<210>   112
<211>   10
<212>   DNA
<213>   Homo sapiens

<400>   112
tcctgcctgt   10

<210>   113
<211>   10
<212>   DNA
<213>   Homo sapiens

<400>   113
gcagggaaga   10

<210>   114
<211>   10
<212>   DNA
<213>   Homo sapiens

<400>   114
ccagccacac   10

<210>   115
<211>   10
<212>   DNA
<213>   Homo sapiens

<400>   115
gggcctccag   10

<210>   116
<211>   10
<212>   DNA
<213>   Homo sapiens

<400>   116
agggcctcgg   10

<210>   117
<211>   10

POR.ST25.txt

```
<212>   DNA
<213>   Homo sapiens

<400>   117
ccaccgccac                                                          10


<210>   118
<211>   10
<212>   DNA
<213>   Homo sapiens

<400>   118
gggcctcggt                                                          10


<210>   119
<211>   10
<212>   DNA
<213>   Homo sapiens

<400>   119
tccaccgcca                                                          10


<210>   120
<211>   10
<212>   DNA
<213>   Homo sapiens

<400>   120
agaccaaggc                                                          10


<210>   121
<211>   10
<212>   DNA
<213>   Homo sapiens

<400>   121
tgatgcggcc                                                          10


<210>   122
<211>   10
<212>   DNA
<213>   Homo sapiens

<400>   122
aaggagcctg                                                          10


<210>   123
<211>   10
<212>   DNA
<213>   Homo sapiens

<400>   123
gttctccccg                                                          10


<210>   124
<211>   10
<212>   DNA
<213>   Homo sapiens

<400>   124
gcgctcaccc                                                          10
```

POR.ST25.txt

```
<210>   125
<211>   10
<212>   DNA
<213>   Homo sapiens

<400>   125
ggcaggggcc                                                      10


<210>   126
<211>   10
<212>   DNA
<213>   Homo sapiens

<400>   126
tgccgccgct                                                      10


<210>   127
<211>   10
<212>   DNA
<213>   Homo sapiens

<400>   127
gtcctcatcc                                                      10


<210>   128
<211>   10
<212>   DNA
<213>   Homo sapiens

<400>   128
agggccacag                                                      10


<210>   129
<211>   10
<212>   DNA
<213>   Homo sapiens

<400>   129
gggcactgtg                                                      10


<210>   130
<211>   10
<212>   DNA
<213>   Homo sapiens

<400>   130
gtctgtgggt                                                      10


<210>   131
<211>   10
<212>   DNA
<213>   Homo sapiens

<400>   131
cccactcact                                                      10


<210>   132
<211>   10
<212>   DNA
```

POR.ST25.txt

```
<213>  Homo sapiens

<400>  132
taatcagctc                                                              10


<210>  133
<211>  10
<212>  DNA
<213>  Homo sapiens

<400>  133
gggagccagg                                                              10


<210>  134
<211>  3120
<212>  DNA
<213>  Homo sapiens

<220>
<221>  allele
<222>  (30)..(30)
<223>  PS1: polymorphic base G or A


<220>
<221>  misc_feature
<222>  (61)..(120)
<223>  n's represent sequence between PS1 and PS2


<220>
<221>  allele
<222>  (150)..(150)
<223>  PS2: polymorphic base C or T


<220>
<221>  misc_feature
<222>  (181)..(240)
<223>  n's represent sequence between PS2 and PS3


<220>
<221>  allele
<222>  (270)..(270)
<223>  PS3: polymorphic base C or T


<220>
<221>  misc_feature
<222>  (301)..(360)
<223>  n's represent sequence between PS3 and PS4


<220>
<221>  allele
<222>  (390)..(390)
<223>  PS4: polymorphic base C or T


<220>
<221>  misc_feature
<222>  (421)..(480)
<223>  n's represent sequence between PS4 and PS5
```

Page 33

POR.ST25.txt

```
<220>
<221>  allele
<222>  (510)..(510)
<223>  PS5: polymorphic base G or A


<220>
<221>  misc_feature
<222>  (541)..(600)
<223>  n's represent sequence between PS5 and PS6


<220>
<221>  allele
<222>  (630)..(630)
<223>  PS6: polymorphic base C or T


<220>
<221>  misc_feature
<222>  (661)..(720)
<223>  n's represent sequence between PS6 and PS7


<220>
<221>  allele
<222>  (750)..(750)
<223>  PS7: polymorphic base C or T


<220>
<221>  misc_feature
<222>  (781)..(840)
<223>  n's represent sequence between PS7 and PS8


<220>
<221>  allele
<222>  (870)..(870)
<223>  PS8: polymorphic base A or G


<220>
<221>  misc_feature
<222>  (901)..(960)
<223>  n's represent sequence between PS8 and PS9


<220>
<221>  allele
<222>  (990)..(990)
<223>  PS9: polymorphic base C or T


<220>
<221>  misc_feature
<222>  (1021)..(1080)
<223>  n's represent sequence between PS9 and PS10


<220>
<221>  allele
<222>  (1110)..(1110)
<223>  PS10: polymorphic base C or G


<220>
<221>  misc_feature
```

POR.ST25.txt

```
<222>   (1141)..(1200) .
<223>   n's represent sequence between PS10 and PS11


<220>
<221>   allele
<222>   (1230)..(1230)
<223>   PS11: polymorphic base G or A


<220>
<221>   misc_feature
<222>   (1261)..(1320)
<223>   n's represent sequence between PS11 and PS12


<220>
<221>   allele
<222>   (1350)..(1350)
<223>   PS12: polymorphic base C or T


<220>
<221>   misc_feature
<222>   (1381)..(1440)
<223>   n's represent sequence between PS12 and PS13


<220>
<221>   allele
<222>   (1470)..(1470)
<223>   PS13: polymorphic base C or T


<220>
<221>   misc_feature
<222>   (1501)..(1560)
<223>   n's represent sequence between PS13 and PS14


<220>
<221>   allele
<222>   (1590)..(1590)
<223>   PS14: polymorphic base C or T


<220>
<221>   misc_feature
<222>   (1621)..(1680)
<223>   n's represent sequence between PS14 and PS15


<220>
<221>   allele
<222>   (1710)..(1710)
<223>   PS15: polymorphic base T or C


<220>
<221>   misc_feature
<222>   (1741)..(1800)
<223>   n's represent sequence between PS15 and PS16


<220>
<221>   allele
<222>   (1830)..(1830)
<223>   PS16: polymorphic base C or G
```

POR.ST25.txt

```
<220>
<221> misc_feature
<222> (1861)..(1920)
<223> n's represent sequence between PS16 and PS17


<220>
<221> allele
<222> (1950)..(1950)
<223> PS17: polymorphic base G or A


<220>
<221> misc_feature
<222> (1981)..(2040)
<223> n's represent sequence between PS17 and PS18


<220>
<221> allele
<222> (2070)..(2070)
<223> PS18: polymorphic base T or C


<220>
<221> misc_feature
<222> (2101)..(2160)
<223> n's represent sequence between PS18 and PS19


<220>
<221> allele
<222> (2190)..(2190)
<223> PS19: polymorphic base G or T


<220>
<221> misc_feature
<222> (2221)..(2280)
<223> n's represent sequence between PS19 and PS20


<220>
<221> allele
<222> (2310)..(2310)
<223> PS20: polymorphic base C or T


<220>
<221> misc_feature
<222> (2341)..(2400)
<223> n's represent sequence between PS20 and PS21


<220>
<221> allele
<222> (2430)..(2430)
<223> PS21: polymorphic base T or C


<220>
<221> misc_feature
<222> (2461)..(2520)
<223> n's represent sequence between PS21 and PS22
```

POR.ST25.txt

```
<220>
<221>  allele
<222>  (2550)..(2550)
<223>  PS22: polymorphic base C or T


<220>
<221>  misc_feature
<222>  (2581)..(2640)
<223>  n's represent sequence between PS22 and PS23


<220>
<221>  allele
<222>  (2670)..(2670)
<223>  PS23: polymorphic base C or G


<220>
<221>  misc_feature
<222>  (2701)..(2760)
<223>  n's represent sequence between PS23 and PS24


<220>
<221>  allele
<222>  (2790)..(2790)
<223>  PS24: polymorphic base C or G


<220>
<221>  misc_feature
<222>  (2821)..(2880)
<223>  n's represent sequence between PS24 and PS25


<220>
<221>  allele
<222>  (2910)..(2910)
<223>  PS25: polymorphic base G or A


<220>
<221>  misc_feature
<222>  (2941)..(3000)
<223>  n's represent sequence between PS25 and PS26


<220>
<221>  allele
<222>  (3030)..(3030)
<223>  PS26: polymorphic base T or C


<220>
<221>  misc_feature
<222>  (3061)..(3120)
<223>  n's represent sequence 3' to PS26


<400>  134
atcaagtcta gcagacggct ggcgtcaccr catagaggat gtcccggtgg ccaggagagg    60
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   120
catcgtgttc tacggctccc agacggggay tgcagaggag tttgccaacc gcctgtccaa   180
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   240
ctgtagggcg cccctcagca gggggaggcy ggcagggagt ggggtcctgg gaagacgtcc   300
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   360
ggcgcggtcc tgtccctgtt tctgcaggcy gacctgagca gcctgccaga gatcgacaac   420
```

POR.ST25.txt

```
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    480
gagctttggg gatggggtgg ggtcggggcr tgcctggcac caggtaccgt tgccacatgg    540
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    600
cttcatcacc tggcgagagc agttctggcy ggccgtgtgt gaacactttg gggtggaagc    660
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    720
atcacctggc gagagcagtt ctggccggcy gtgtgtgaac actttggggt ggaagccact    780
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    840
agggagcgtg gcttggggca gacggctctr tggccactgg tgcaccccag gctcagtctg    900
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    960
gccactggtg caccccaggc tcagtctgcy gtgtatcccc atatccccac agggcccttc   1020
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   1080
agattccctg tgctttgtgc aaccagaags gtccttggag acggagactc agatcaaagc   1140
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   1200
tttgtgcaac cagaagcgtc cttggagacr gagactcaga tcaaagcccc ggccgctcac   1260
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   1320
cttggagacg gagactcaga tcaaagcccc ggccgctcac tgtgcttctc tcctccccac   1380
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   1440
tttccagcac cagctggggc acctgttgcy gcagagctgg cccaaggtgt caccccctcc   1500
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   1560
gcagagctgg cccaaggtgt caccccctcy tgccgcagcc acccatcccc aggaggcggc   1620
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   1680
cagagctggc ccaaggtgtc accccctccy gccgcagcca cccatcccca ggaggcggcc   1740
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   1800
ggcggccgcc taccccaagt cctgcctgts tcttccctgc agaggagtcc aacaagaagc   1860
nnnnnnnnnn nnnnnnnnnn.nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   1920
gagggccggc actgccctgc cagccacacr ctggaggccc agccctgctc acagcaggca   1980
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   2040
ttggggccgg ggctgggcaa gggcctcggy gtggcggtgg agctcacacg gccctcccca   2100
nnnnnnnnnn.nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   2160
tggggccggg gctgggcaag ggcctcggtk tggcggtgga gctcacacgg ccctccccac   2220
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   2280
gcggtggttg tggagtacga gaccaaggcy ggccgcatca acaagggcgt.ggccaccaac   2340
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   2400
caccaactgg ctgcgggcca aggagcctgy cggggagaac ggcggccgtg cgctggtgcc   2460
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   2520
ggtgacgact gggagccccg cgctcacccy ggccctgcc acgcaggcaa ggaggtgggg   2580
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   2640
gacgctgctg tactacggct gccgccgcts ggatgaggac tacctgtacc gggaggagct   2700
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   2760
aggtgggcat gaggctggca gggcacacags cacagtgccc ccctcacagc accacccttg   2820
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   2880
aggcggtgcc cacatctacg tctgtgggtr agtgagtggg gtcactggaa taggggggcag   2940
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   3000
ccccacagac tccggcctgt aatcagctcy cctggctccc tcccgtagtc tcctgggtgt   3060
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn   3120
```

Page 38